AD-A047 299 AMERICAN INSTITUTES FOR RESEARCH WASHINGTON D C F/G 5/9
AN INFORMATION PROCESSING APPROACH TO PERFORMANCE ASSESSMENT. I--ETC(U)
NOV 77 A M ROSE, K FERNANDES N00014-76-C-0871
UNCLASSIFIED AIR-58500-TR NL

| OF |
AD
A047 299

END
DATE
FILMED
1 - 78
DDC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 An Information Processing Approach to Performance Assessment. I. Experimental Investigation of an Information Processing Performance Battery.		5. TYPE OF REPORT & PERIOD COVERED 9 Technical Report, 12 1 Apri[redacted]7 — 15 Nov[redacted]77,
		6. PERFORMING ORG. REPORT NUMBER 14 AIR-58500-TR, TR-1
7. AUTHOR(s) A. M. Rose K. Fernandes		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-76-C-0871 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS AMERICAN INSTITUTES FOR RESEARCH 1055 Thomas Jefferson Street, N.W. Washington, D.C. 20007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 17 61153N; RR042-04-01 16 RR042-04 NR 150-391
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs, Office of Naval Research - Code 458 Arlington, Virginia 22217		12. REPORT DATE 11 25 Nov[redacted] 1977
		13. NUMBER OF PAGES 96
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 10 Andrew M. Rose Kathleen Fernandes 12 95p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Personnel Technology
Individual Differences
Personnel Assessment
Information Processing
Cognitive Processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the first study in a program of research dealing with the development and validation of a comprehensive standardized test battery that can be used as an assessment device for the evaluation of performance in a wide variety of situations. The standardized battery is being designed to possess high reliability and predictive validity for a wide variety of criterion tasks. next

DD FORM 1 Jan 73 1473 023 450

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

LB

20. Abstract (Cont'd.)

Equally important, the battery is being designed to include tests that possess construct validity: there will be a firm theoretical and empirical base for inferring the information possessing structures and functions that the tests purport to measure. It is expected that such a battery will permit improved personnel management decisions to be made for a wider variety of Navy-relevant jobs than is currently possible using existing techniques.

The major purpose of the present experimental study was to determine properties of the tasks selected for inclusion in the test battery. The primary questions addressed during this phase concerned the replicability of previous findings and the adequacy of the tests to provide measures of individual differences. In addition information concerning the construct validity of the tasks and population norms for the resultant measures could be investigated. With the relatively large data base employed (54 subjects), additional data concerning the ability of the set of measures to separate individuals within the population could be examined.

The tests investigated included the following:

Letter classification (Posner and Mitchell)
Lexical decision making (Meyer)
Graphemic and phonemic analyses (Bar
Short term memory scanning (Sternberg)
Memory scanning for words and categories (Juola)
Linguistic verification (Clark and Chase)
Recognition memory (Shepard and Teghtsoonian)
Semantic memory retrieval (Collins and Quillian)

Several questions were addressed in this phase of the research. First, replicability of previous experimental work with similar paradigms was investigated. In general, the results were quite compatible with previous findings for all eight tasks. The second area addressed concerned the establishment of the reliability, validity, and independence of the tasks being studied. In general the reliabilities for most measures was quite high ($r \geq .50$). The measures were also analyzed to determine practice effects and the character of the response distributions in the population for each of the measures.

In order to address validity-type issues, inter- and intra- task correlations were calculated. In general, these analyses support the construct validity of the tasks and measures.

AIR-58500-TR

An Information Processing Approach to Performance Assessment:

I. Experimental Investigation of an Information Processing Performance Battery

ANDREW M. ROSE

KATHLEEN FERNANDES

TECHNICAL REPORT

November 1977

Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, Arlington, Virginia

Contract No. N00014-76-C-0871

AMERICAN INSTITUTES FOR RESEARCH / 1055 Thomas Jefferson Street, NW, Washington, DC 20007

ACKNOWLEDGMENTS

The authors wish to express their sincere appreciation to the Army Research Institute for their kind permission to allow us use of their computer laboratory facilities. In particular, the assistance of the following ARI staff was vital to the conduct of our experiment: Dr. Abraham H. Birnbaum, Assistant to the Director of Organizations and Systems Research Laboratory (OSRL): Mr. John R. Schjelderup, Mr. Joseph P. Severo, Mr. Arthur Lynch, and Mr. Charles Marshall, all of the Research Support Group under OSRL; and Mr. Donald E. Andre, Computer Operations Supervisor of the data processing center.

Also, we would like to acknowledge the assistance of Dr. David E. Meyer, Dr. James F. Juola, Dr. Jonathan Baron, and Dr. Allan M. Collins, who provided us with stimulus materials, instructions, and other materials useful in the conduct of our study.

Finally, we would like to acknowledge the contribution of Dr. Paul W. Fingerman of AIR, who was the principal author and constructor of the computer programs used in the real-time conduct of the experiment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In order to improve the accuracy of personnel selection, classification, guidance, and the design of training programs, procedures for making meaningful evaluations or differentiations among people with respect to relevant characteristics must be established. For example, different occupations and professions place differential demands upon human cognitive abilities. It follows, therefore, that maximization of the fit between people and jobs necessitates both the specification of essential requirements for each type of work and the assessment of personnel with respect to the characteristics needed on the several jobs. As a consequence the central question of manpower utilization revolves around performance assessment and individual differences: What are the performance requirements for a particular task or job, and how can the person be found or trained who best matches those requirements?

Almost since its inception approximately one hundred years ago one of the major concerns of scientific psychology has been the assessment and evaluation of individuals along a wide variety of dimensions. In fact, despite their controversial status in recent times, it can be argued that personnel and evaluation tests are the most widespread contributions of the science of psychology, and possibly the most important practical utilizations of empirically derived information concerning human performance.

However, with the growing maturity of empirical psychology, personnel tests and other similar tests have been the source of a certain amount of dissatisfaction, particularly among experimental psychologists. Much of this dissatisfaction can be directly attributed to the unanalytical (with respect to underlying and more basic kinds of performance) nature of most personnel tests. That is, from the perspective of many, most notably those subscribing to an information-processing theoretical framework, the types of tasks represented in global tests such as those purporting to measure "job performance" too frequently are either extremely difficult to categorize as to precisely what is being measured, or else involve a not easily interpretable mixture of two or more identifiable processes or stages.

An initial attempt to develop a more analytical and incisive battery of information-processing performance tests has been reported by Rose (1974). Rose compiled a series of tests designed to be useful in assessing individual differences in a wide variety of information-processing skills and capacities. Some of the criteria implicit in the initial selection of tests for this Information Processing Performance Battery (IPPB) were:

(1) the tests should measure specific processes;

(2) the tests should be sensitive and yet short and easy to administer;

(3) the tests should yield reliable measures of performance;

(4) the tests should be statistically independent;

(5) the tests should measure basic abilities and hence must be relatively abstract to minimize effects of prior experience; and

(6) taken as a whole, the tests in the battery should not be "criterion-based", i.e., the tests should not be selected with the aim of predicting one particular type of job or activity.

After carefully selecting a series of nine tests that satisfied these criteria, Rose administered the IPPB for three sessions each to 100 college-age subjects -- 50 males and 50 females. Extensive correlational analyses were conducted on the data resulting from these administrations to determine the degree of relationship among the various tests and the reliabilities of each test.

This work has been extended in a current ONR-sponsored research program being conducted at the American Institutes for Research. The basic objective of this work is to further develop and validate the IPPB so that eventually it can be used as an assessment device for the evaluation of performance in a wide variety of situations. However, at the initiation of the present project, the IPPB tapped only a limited number of information-processing functions and structures. In order to increase the scope of the

functions represented, it was considered necessary to refine and expand the existing tests in the battery. The general strategy employed in the development of the IPPB is described in the following paragraphs.

The relevant literature of the past several years was carefully and extensively reviewed in order to locate major omissions as a basis for major revision of the IPPB. Several methods were used. First, "standard" computer-based search and retrieval systems such as PASAR and MEDLARS were employed to identify relevant references. Second, a number of recently published texts on information processing were reviewed, as were books of readings on memory and cognition. Given the growth of literature relating to information processing during the past several years, a review of selected secondary sources was considered to be an economical way of obtaining an overview of the research that had been accomplished since the IPPB was developed. A third procedure was to collect references dealing with particular information-processing constructs, processes, or tasks that were not represented previously. Toward this end a narrow-focus, topic-specific literature search was initiated.

In using the three literature identification procedures, special consideration was given to major areas of information processing in which a large body of research has accumulated since the IPPB was developed. These areas included memory, psycholinguistics, and visual information processing. Within each of these areas, relevant paradigms and constructs were further evaluated with several criteria in mind:

1. The information-processing construct or concept had to have a history of empirical and/or theoretical support. The interest here was in constructs that had been developed over a period of time and in research paradigms that had been replicated under a variety of conditions. This criterion was relaxed only in instances where a paradigm was considered to be a "classic" measure of a particular construct but where no evidence of replication could be found in the literature.

2. There had to be an adequate theoretical rationale for the paradigm actually measuring the particular information-processing construct that it

was intended to measure. The focus was on construct validity rather than theoretical sophistication. Studies concerned primarily with the development of mathematical models for certain operations, with the task itself of only ancillary relevance, were excluded from further consideration.

3. The experimental task itself had to be one that was adaptable to a paper-and-pencil format, to a small digital computer, or to some other form that could be easily administered in a group setting.

4. Enough performance data had to be available so that preliminary estimates could be made regarding the extent of individual variation expected for the task.

The result of the screening activity was a set of 15 tasks that seemed to be prime candidates for inclusion in the IPPB. These tasks were adapted or modified into practical formats, and the methodological refinements were evaluated in a series of in-house, informal pilot studies to determine the feasibility of alternative adaptations of tasks, instructions, stimuli, and timing. At the completion of these studies, all of the tasks had been evaluated to determine their empirical and theoretical support, logical feasibility, reliability, and, to a limited degree, their construct validity. As a result of this evaluation, eight tasks were retained and considered worthy of more extensive experimental investigation.

The next phase of the present project was to determine the properties of the tasks when they were assembled into a research battery. The primary questions addressed during this phase concerned the replicability of previous findings and the adequacy of the tasks to provide measures of individual differences. In addition, information concerning the construct validity of the tasks and sample norms for the resultant measures could be investigated. With the relatively large data base employed, additional issues concerning the ability of the set of measures to separate individuals within the population could be examined.

The next sections provide a detailed description of the conduct of the experiment designed to fulfill those purposes. First, the eight tasks included in the experiment are described. This presentation is followed by a description of task measures, experimental method, and results.

General Task Overview

During the course of task selection for the first phases of this project, we have attempted to remain unbiased toward any general or specific information-processing model or theory. It has not been our goal to develop a general model of human information processing, nor has our task selection been organized around or directed towards a new "structure of intellect." In this regard, we echo the opinions expressed by Melton (1967, p. 241):

> We have at this time no general theory of human learning and performance. Therefore, we have no necessary and sufficient list of process constructs or variables that can serve as the foci of individual-differences research .... There is no magic whereby the processes that should be examined with respect to individual differences can be identified. The process concepts to be examined will depend on the level of analysis that our theoretical-experimental approach has achieved and on the level of analysis and range of task variables that the theoretical model attempts to encompass. (p. 241 ff)

Nevertheless, as a result of our task selection criteria and pilot task-selection efforts, a post-hoc organizational structure has manifested itself. This structure is conceptually useful in delimiting the scope of tasks included in the IPPB as well as in providing a basis upon which construct validity issues can be discussed.

Carroll (1974), in an initial attempt to characterize a set of factor-analytically-derived abilities in terms of cognitive processes, developed an organizational system of information-processing "operations" that is conceptually similar to the simple structure presented here. He defines <u>operations</u> as "control processes that are explicitly specified, or implied, in the task instructions ... and that must be performed if the task is to be successfully completed." These operations are of three types: attentional, memorial, and executive. Of particular interest are the latter two, which he further subdivided as follows: There are three kinds of memorial operations: <u>storing</u>, <u>searching</u>, and <u>retrieving</u>. Executive operations are ex-

emplified by such things as: simple judgments of stimulus attributes such as to reveal identity, similarity, or comparison between two stimuli; manipulations of memorial contents, such as "mentally rotating" a visuospatial configuration; and information transformations which produce "new" elements from combinations, reductions, etc., of old elements. The second dimension of this organization is what he calls "modality or contents" -- basically, the form in which the stimuli are processed by the operations. While Carroll's organization was developed in a different context than the present project, it is a good illustration of the level of analysis that we will continue to use. In fact, we will retain several of the above distinctions.

In general, all tasks included in the present study can be described as a series of operations, where an operation is defined as above. Each task can be specified by some combination of eight such operations. These operations are described below.

Encoding: the operation by which information is input into the system, and including the initial set of processes that converts the physical stimulus to a form which is "appropriate" for the task. Different task demands may require different levels of analysis of the stimulus. Posner (1969) has called this dimension "abstraction" -- the process by which different types of information about the stimulus are extracted; in other words, the level of stimulus analysis demanded by the task. For example, a visual search task might require only that the subject extract physical or structural information about the stimulus, a memory search task might require the extraction of name information, and a semantic search task might necessitate semantic or "meaning" information.

Constructing: the operation by which new information structures are generated from information already in the system. This is what Neisser (1967) and others have called "synthesis"; in the present context, we will limit the use to situations where additional features of the stimuli must be abstracted, beyond those initially encoded.

Transforming: the operation by which a given information structure is converted into an equivalent structure necessary for task performance. In contrast to constructing, transformations do not involve any new information

abstraction; rather, this operation requires the application of some stored rules to the information structure already present.

Storing: the operation by which new information is incorporated into existing information structures, while its entire content is retained.

Retrieving: the operation by which previously stored information is made available to the processing system.

Searching: the operation by which an information structure is examined for the presence or absence of one or more properties. The information structure examined may be one already in the processing system or one external to it (e.g., a visual array).

Comparing: the operation by which two information structures (again, either internal or external to the processing system) are judged to be the same or different. The information structures need not both be physical entities (as in the comparison of two objects); likewise, a physical entity can be compared to a stored representation or description in order to determine identity.

Responding: the operation by which the appropriate (motor) action is selected and executed. In many information-processing investigations, the response operation is itself the object of study. Various microprocesses have been uncovered; however, the current study was designed to minimize performance variability due to differential response demands of the tasks.

Each of the eight tasks included in the present study will be described first, in terminology employed by the particular investigators, and second, as a function of some of the operations just elucidated. These latter descriptions do not, of course, represent the original authors' conceptions of the paradigms.

Task Descriptions*

Letter classification (Posner task). The process of matching or recognition at various levels of stimulus complexity is basic to most cognitive

* For ease of discussion in this section and in following sections, a shortened label for each of these tasks will be used, namely the principal author's name. Thus, the letter classification task will be addressed as the "Posner" task, etc.

tasks. Posner and Mitchell (1967) developed an experimental paradigm which "provides an opportunity to observe processing at different levels within the experiment." In their task, the subject was shown pairs of letters and had to decide whether the letters were the same or different. The independent variable was the instruction upon which the subject was told to make the classification. The instructions used to define "same" were:

1. Physical identity (e.g., the pair AA is to be classified as "same" while AB is "different"); or

2. Name identity (e.g., Aa is "same", Ab is "different"), or

3. Category or rule identity (e.g., if the rule is one of letter category, a stimulus pair is "same" if both members are vowels or if both members are consonants, such as AE or BD).

The typical findings were that the classification reaction times (RTs) increased as the instructions varied in the above order. The ordinal relationships among these processing "nodes" (and the time differences between them) were quite reliable and have been demonstrated to generalize to other stimuli (e.g., numbers, Gibson figures). It also has been shown that these types of classifications are serial (i.e., subjects derive the name of the letters before proceeding to analyze whether they are both vowels or both consonants).

The procedure in the current project used the three Posner and Mitchell experimental conditions. The category identity condition was modified slightly in that the vowel category and the consonant category were tested conjointly. In the Posner and Mitchell study, the two categories were tested in separate blocks of trials.

In terms of our operations, the three conditions all involve an initial encoding of letters. In the physical match case, the subject then compares the representations of the letter patterns, and finally selects and executes the appropriate response. The name case requires an additional operation of retrieval of "name" information from long-term memory (LTM). Subjects then

compare the letter names. In the rule condition, experimental evidence indicates that subjects retrieve name information prior to retrieving categorical information (i.e., subjects categorize names of letters as vowels or consonants rather than the physical patterns). They then compare the representations of the letter categories.

Lexical decision making (Meyer task). Rubenstein, Garfield, and Millikan (1970) developed a procedure designed to investigate the processes by which humans can recognize written words. On each trial in their paradigm, a string of letters was presented and the subject had to judge whether it was an English word or nonword. Performance on this lexical decision task depended on operations that mediated the recognition of printed words in various contexts -- that is, graphemic and/or phonemic encoding, followed by accessing of lexical memory. Various investigators have argued that printed words are recognized directly from visual representations (graphemes), while others have claimed that recognition is mediated by a phonological (phonemic) representation.

The Rubenstein et al. procedure has been modified by Meyer (e.g., Meyer, Schvaneveldt, and Ruddy, 1974) in order to separate the effects of graphemic and phonemic factors on recognition. As in the Rubenstein et al. experiments, subjects were presented with two strings of letters, displayed successively, on each trial. Reaction time (RT) was measured for each string separately. The critical variables were the graphemic and phonemic relations within the pairs of words. For example, the words could be both graphemically and phonemically similar (e.g., BRIBE - TRIBE), graphemically similar but phonemically dissimilar because they do not rhyme (e.g., COUCH - TOUCH), and so on. Meyer et al. formulated and tested various hypotheses concerning the relative speed of recognition for word pairs; for example, it was found that graphemic similarity alone inhibited performance (e.g., in the pair COUCH-TOUCH, RT to TOUCH was slower than predicted from baseline control conditions). In contrast, phonemic as well as graphemic similarity facilitated recognition (e.g., in the pair BRIBE-TRIBE, RT to TRIBE was faster than to the second word of graphemically and phonemically dissimilar word pairs).

The Meyer et al. paradigm was modified in the present study to include a category of phonemically similar but graphemically dissimilar word pairs (e.g., LAUGH-HALF).

In terms of operations, both the "word" and "nonword" stimulus presentations require the subject to encode a letter string; the "mediation" hypothesis can be formulated as the (optional) construction of a phonemic or graphemic representation. Presumably, this paradigm will identify those subjects who have a propensity for one construction or the other. Following this construction, both conditions require the search of LTM for a match. When a match is found (in the word case), subjects select and execute the appropriate response. If a match is not found (in the nonword case), it is hypothesized that subjects conduct a further search -- this time, of the lexical rules in LTM in order to decide whether or not a letter string is an acceptable construction. Following this search, subjects select and execute the appropriate response.

Graphemic and phonemic analysis (Baron task). Baron (1973; Baron & McKillop, 1975) has developed a procedure for the study of individual differences in the speed of phonemic (acoustical) and graphemic (visual) analysis of printed information (e.g., sentences or phrases). He argued that lexical memory can be accessed through both visual and phonological representations of a printed word; also, he argued that a visual analysis is the faster of the two for normal readers. The basic paradigm used in his studies was to "force" subjects to analyze phrases visually and phonologically. More specifically, he had subjects decide whether various printed phrases made sense or were nonsense. Three conditions were required. In the first condition two kinds of phrases were used: sense (S) phrases, and those which sounded sensible because of a homophone (e.g., IT'S KNOT SO) but looked like nonsense (called H phrases). In this first condition (SH), subjects were instructed to classify a phrase as making sense or nonsense on the basis of its appearance (so that H phrases were judged as nonsense). The second condition used H phrases and true nonsense (N) phrases (e.g., NEW I CAN'T). In this second condition (HN), subjects were instructed to classify the phrases on the basis of how they sounded, so that H phrases were judged as making sense. The third

condition used S and N phrases. In this third condition (SN) subjects were free to choose whatever basis they preferred for making S and N judgments. The basic analysis was to determine which of the first two conditions better predicted the third condition. For example, if a particular subject was a "visual" encoder, he should have had "problems" with the HN condition and his SH performance should have been a good predictor of SN performance. The results (as reported in Baron and McKillop, 1975) indicated the existence of reliable and predictable individual differences: some subjects were "visual", others "phonemic" encoders.

The procedure in the current study used Baron's three conditions. However, we obtained RTs on a trial-by-trial basis rather than after a trial block.

In terms of operations, all three conditions (SH, HN, and SN) require the subjects to encode semantic phrases. Following this encoding, the different conditions force subjects to construct either visual or acoustic representations of each phrase: the SH condition needs a visual representation, the HN condition needs an acoustic representation, and the SN requires either an acoustic or a visual representation. Following this construction, subjects must search LTM for what we will call "phrase rules" -- that set of information or rules that enables them to decide whether or not a phrase meets acceptable language structure rules. In all conditions, subjects then select and execute the appropriate response.

Short-term memory scanning (Sternberg task). Sternberg (1967, 1969) developed an experimental paradigm to "study the ways in which information is retrieved from memory when learning and retention are essentially perfect" (Sternberg, 1969, p. 423). The general procedure was to present a list of items for memorization that was short enough to be within the immediate memory span (typically, this "memory set" contained 1-4 items). Next, the subject was asked a question about the memorized list (again, typically, the question concerned the presence or absence of a stimulus from the memorized set), and his delay in responding to the question was measured. The particular manifestation of this general procedure used in the current work was the "item-

recognition task." The stimulus ensemble consisted of the digits 1 through 9. On each trial, a set of digits was selected arbitrarily and was defined as the positive or memory set. After a short pause, a test stimulus (a single digit) was presented. The subject had to decide whether the test digit was a member of the positive set. Performance was measured in terms of RT from test-stimulus onset to response.

The typical findings were that the functions relating RT to memory set size are approximately linear, and with roughly equal slopes for positive and negative responses. This outcome has been observed in many different situations, including differences in stimulus ensemble, subject group differences, and memory set sizes. The paradigm also resulted in reliable individual differences with respect to the slope and intercept parameters of the RT by memory-set-size function. The procedure in the current study was essentially a replication of Sternberg's "varied set" procedure, wherein the memory set was changed from trial to trial.

In terms of operations, we will consider those events that take place starting from the presentation of the target number, since it is assumed that this paradigm does not measure any aspect of storing or retrieving information. Thus, when the target stimulus is presented, subjects must encode the number. Following this encoding (which may be visual or acoustic, depending upon the nature of the representation of the memory set), subjects compare the target with the memory set. This comparison is (generally) accomplished in a serial, exhaustive manner -- all items in the memory set are compared prior to the selection and execution of the appropriate response.

Memory scanning for words and categories (Juola task). Memory search processes for word names and for categorical information about words were investigated in an experiment by Juola and Atkinson (1971). They used a short-term memory search paradigm similar to that used by Sternberg (1967) in which a short list of items was presented, followed by a single probe item that might or might not be a member of the memorized list. Two major conditions were run in the Juola and Atkinson study: a "word scan" condition and a "category scan" condition. In the first condition, the memory

set consisted of from one to four different words. A positive probe stimulus was one of the words in the memorized list, while a negative probe was a word that did not match any of the memory set words. Thus, this condition was essentially a replication of the Sternberg paradigm using words rather than numbers. The second condition in the Juola and Atkinson study also involved a memory set of from one to four words; however, these words were semantic category labels (e.g., COLOR, RELATIVE, etc.). Positive probe stimuli were instances of one of the memory set categories (e.g., if the memory set was COLOR, RELATIVE, a positive probe might be BLUE).

The results of this experiment (and a replication by Juola and McDermott, 1976) showed an increase in response time with the number of memory set items in both conditions. Furthermore, when linear functions were fit to the data, the functions had equivalent intercepts for the two conditions, but the slope was much greater for the categorization condition. The authors argued that the comparability of intercepts indicated that categorization and comparison involve many similar processes that do not depend upon the size of the memory set (e.g., probe word encoding, response decision and execution), while a difference in slope indicated that fundamentally different types of search or comparison processes are involved in the two conditions.

The procedure used in the present research was a modification of the Juola and Atkinson task in that: (1) the same category labels were used in both conditions; (2) a relatively small set of categories was employed; (3) several exemplars of each category were used in the categorization condition; and (4) negative probes were members of other categories used as memory set items.

In terms of operations, the "word scan" condition is essentially equivalent to the Sternberg task in that it requires the encoding of the target stimulus (in this case, a word rather than a number), followed by a serial comparison of that representation with the items in the memory set, and the selection and execution of the appropriate response. The "category scan" condition requires an additional operation -- the retrieval of categorical information from LTM. The pattern of their results indicates that this

retrieval operation is performed each time the target word is compared to a member of the memory set, rather than just once. The results also suggest that the comparison operation is serial and self-terminating in this condition, rather than exhaustive.

Linguistic verification (Clark task). Clark and Chase (1972) developed and tested a model to account for how people compare information from linguistic and pictorial sources. Their model applied to a particular type of sentence verification task in which the subject was presented with a display containing a sentence and a picture. The sentence was of the form "star (plus) is (is not) above (below) plus (star)" and the picture was either $\overset{*}{+}$ or $\overset{+}{*}$. The subject had to decide whether the sentence was a true or false description of the picture. The model accounted for the latencies of the subject's judgments in terms of four operations or stages (sentence encoding, picture encoding, comparing, and responding) which were serially ordered, with component latencies that were additive. The subject formed internal representations of the sentence and the picture in terms of their underlying propositions and then performed a series of comparison operations to check the overall congruence of the representations. Clark and Chase found that verification time consisted of the addition of one or more of four parameters that accounted for 99.8 percent of the variance in response latencies.

The procedure in the current study was a replication of the sentence verification task as used by Clark and Chase.

The above description is compatible with the operations terminology used here in that each type of sentence requires an encoding of a sentence and a picture, a comparison of those representations, and a response selection and execution. In addition, our terminology requires that two additional operations be included: constructing of what has been called a "kernel" representation, and transforming of the representation, based on the particular modifiers in the various sentence types. For example, Clark and Chase argue that "below" is transformed into "not above"; similarly, they argue that negations and "truth indices" are likewise transformed, depending upon the given sentence configuration.

Semantic memory retrieval (Collins and Quillian task). A topic of considerable concern to psychologists is how semantic information is stored, organized, and retrieved. Of the many paradigms used to investigate these issues, one of particular interest requires subjects to make true-false decisions about propositions (Collins and Quillian, 1969). Subjects were presented with sentences such as, "A canary can fly." or, "A canary is an animal." and were asked to ascertain the truth of the statement. The results of the Collins and Quillian (1969) studies using this paradigm supported a theory that semantic information is hierarchically organized in memory. Names of semantic categories are stored at the nodes of a network, along with "pointers" that indicate the relationship between that category and others (e.g., subset or superset relationships are represented as a direction to a different, lower- or higher- order node), and "pointers" to other words indicating properties of that category. Given this structural model and a number of assumptions, the authors were able to make predictions about retrieval time. These assumptions are: first, that both retrieving a property from a node and moving up a level in a hierarchy take a measurable amount of time; second, that the times for these two processes are additive wherever one step is dependent on the completion of another step; and third, that the time to retrieve a property from a node is independent of the level of the node.

Collins and Quillian (1969) reported results consistent with hypotheses generated from their model. For example, they found that subjects could confirm sentences such as "A canary is a bird." more rapidly than "A canary is an animal."; furthermore, "property" sentences such as "A canary can sing." were more quickly confirmed than, "A canary has skin.". The former comparison was predicted from the hypotheses that "canaries are a subset of "birds" which are a subset of "animals"; in order to judge that canaries are animals the subject must first access the "bird" node, then the "animal" node. Similar reasoning applies to the second example: "singing" is a property of canaries, while "having skin" is a property of animals.

Subsequent research has generated other storage and retrieval models that could account for these data. However, it was felt that this paradigm was still useful as a means of generating reliable data on how subjects access

a particular (restricted) information structure, especially a structure which could conceivably be organized hierarchically. Hence, the Collins and Quillian paradigm was adapted for the purposes of the current project but interpreted only in terms of the information structures contained in the stimuli. The adaptation involved creating additional sets of positive sentences and generating companion sets of negative sentences according to the property and set rules used with the positive sentences.

In terms of operations, both "superset" and "property" sentences require the subject to encode the sentences and construct kernel representations. Both sentences also require the retrieval of superset information from LTM; "property" sentences also require retrieval of property information. Finally, both sentence types require the selection and execution of the appropriate response.

Recognition memory (Shepard and Teghtsoonian task). Shepard and Teghtsoonian (1961) developed a procedure for measuring the capacity of human memory under "conditions approaching a steady state" - where the possibility of rehearsal is minimized while the interference of preceding material is maximized. They argued that situations which confront people with a continuing sequence of items and require them to retain as much as possible of the most recently presented information (e.g., continuous monitoring of complex displays) involve memory processes differing from those tested by most other paradigms. The procedure they employed was a recognition task: subjects were presented with a lengthy list of items and were asked to identify each item as "old" (i.e., previously presented) or "new". The lists were constructed so that the interlist intervals between the original and test presentations of items varied. The authors were able to infer a retention function for a single item by plotting probability of recognition as a function of test lag.

In addition to standard parameter estimates, this paradigm is ideal for estimating parameters derived from signal-detection theory. Using the observed proportions of the two types of errors (i.e., calling an old item "new" and calling a new item "old"), it is possible to generate, for each

subject, an estimate of d' and beta (respectively, an estimate of "true" discriminability, and the location of the subject's subjective decision bias criterion).

The present experiment used the Shepard and Teghtsoonian procedure; however, the stimuli were reconstructed so that exactly the same number of intervals occurred in a list of items.

In terms of operations, this task requires the encoding and storing (in LTM) of numbers and the retrieval of these numbers from LTM. In addition, the most convenient way to describe the recognition judgment is to consider it as a comparison operation -- subjects compare each number with their LTM set and judge the "strength of activation"; this judgment determines which response will be selected and executed.

Summary

Table 1 presents an overview of the tasks included in the present experiment and the hypothesized operations included in each task condition. The following method section details the specific implementation, procedures, and scoring rules for each task. Instructions for each of the tasks are included in Appendix A.

Table 1
Operational Overview of Tasks

Task Condition	Encode	Construct	Transform	Store	Retrieve	Search	Compare	Respond
	OPERATIONS							
POSNER PHYSICAL	Encode letters						Compare representations of letter patterns	Select and execute response
POSNER NAME	Encode letters				Retrieve name from LTM		Compare representations of letter names	Select and execute response
POSNER CATEGORY	Encode letters				Retrieve name from LTM Retrieve category from LTM		Compare representations of letter categories	Select and execute response
MEYER WORD	Encode letter string	Construct phonemic or graphemic representation				Search in LTM for "word"		Select and execute response
MEYER NONWORD	Encode letter string	Construct phonemic or graphemic representation				Search in LTM for "word" Search in LTM for "word" rules		Select and execute response
BARON SH (Visual)	Encode semantic phrases	Construct visual representation				Search LTM for "phrase" rules		Select and execute response
BARON HN (Acoustic)	Encode semantic phrases	Construct acoustic representation				Search LTM for "phrase" rules		Select and execute response
BARON SN	Encode semantic phrases	Construct visual or acoustic representation				Search LTM for "phrase" rules		Select and execute response
STERNBERG	Encode target number						Compare numbers (serial)	Select and execute response
JUOLA WORD	Encode target number						Compare words (serial)	Select and execute response
JUOLA CATEGORY	Encode target word				Retrieve category from LTM		Compare categories (serial)	Select and execute response
CLARK AND CHASE	Encode sentence Encode picture	Construct kernel representations	Transform "below" representation Transform "negation" representation Transform truth indices				Compare sentence and picture representations	Select and execute response
COLLINS AND QUILLIAN SUPERSET	Encode sentence	Construct kernel representation			Retrieve superset information from LTM			Select and execute response
COLLINS AND QUILLIAN PROPERTY	Encode sentence	Construct kernel representation			Retrieve superset information from LTM Retrieve property information from LTM			Select and execute response
SHEPARD AND TEGHTSOONIAN	Encode numbers			Store items in LTM	Retrieve numbers from LTM		Judge strength of activation	Select and execute response

METHOD

Testing facilities

As the result of an interservice agreement between the Office of Naval Research (ONR) and the Army Research Institute (ARI), the first experiment of the project was conducted at ARI's computer-controlled Information Systems Laboratory. The laboratory was arranged with a specific hardware configuration to accommodate the experiment. Five subject stations, each consisting of a CRT display screen, a typewriter-like keyboard, and a telephone were set up, each in an individual screened off area, but all in close physical proximity to one another. All software (including programs, stimuli, and response buffers) was maintained on a mountable disk pack assigned exclusively to the project.

The software was developed to permit up to four subjects to be run at a time; the fifth station was used by the experimenter to initialize the program (e.g., supply identifying information on each subject, indicate the session number, and so on), and to monitor and control the progress of the experiment. From his station the experimenter could monitor the progress of any individual subject, check on intermediate results, initiate each of the eight individual tasks being studied, or restart the experiment in case of hardware failure. Any station could be used as the experimenter station, and every station could communicate with the selected experimenter station by telephone. In addition, any selected experimenter station was only a few feet away from the other stations so that the experimenter could be on the scene quickly to answer questions, etc.

The software ran each block of trials as a unit, but permitted subjects to proceed through a block at their own pace. Further, instructions both before and after practice blocks could be provided via the CRT screen. Stimuli for each trial were presented on the CRT after preset intervals, and subjects made their responses on the keyboard. Feedback ("correct" or "wrong", and the latency if "correct") was then provided on the CRT screen. Response latencies were timed to the nearest 3 msec or better. Response and latency information was recorded in the computer memory, and written on a disk file at the end of each block of trials. Thus, in case of a system failure, only data from the most recent block of trials, or at

worst the most recent task, would be lost. The disk response files were dumped onto tape after each session for later analysis.

Procedure

The subjects participated in two testing sessions, each approximately two hours in length and scheduled two days apart. All eight tasks were presented in each session, in the same order. Stimulus order in each task was randomized, and where appropriate, different stimuli were used in each testing session.

At the beginning of each session the experimenter conducted a dialogue with the computer to provide information on which station he would use, the number of subjects to be run in the session, which station each would use, the session number (day one or day two), and identifying information on each subject so that his two sessions could later be put together for analysis. He could also indicate that he was executing a restart after a hardware failure, in which case the above information was automatically retrieved. The experimenter then initiated the first task. In this, and every other task, the computer instructed subjects to read the appropriate instructions and call the experimenter if there were any questions. It then waited for a signal from each subject, indicating that he was ready. The computer then provided a block of practice trials; each subject could proceed through this practice block at his own pace. The system then asked if there were any questions and waited again. The experimenter was available by phone or in person to give assistance. After all subjects signaled that they were ready, the system ran each subject through the actual task at his own pace. When all were finished, the experimenter was notified; he could at this point allow a rest break, or start the next task as appropriate.

Subjects

The subjects were 54 female and male students from Georgetown University; they were paid for their participation in the study. Because of occasional computer system failures, the number of subjects administered each task varied slightly.

The following sections describe the detailed procedure for each task.

Posner task

Procedure. Each trial began with two central fixation points, one to the left of center of the display area, the other to the right of center. After a foreperiod delay, both points were replaced simultaneously by letters. The subject judged whether or not the two letters were the same or different. As soon as the subject responded, both letters were removed. The intertrial interval was approximately two seconds. The following schematic shows the events in a typical trial:

	Subject shown letter pattern		Subject responds
physical	AA	→	"same"
	AB	→	"different"
name	aA	→	"same"
	aB	→	"different"
category (vowel)	AE	→	"same"
	AB	→	"different"

Stimuli and Design. Each subject classified pairs of letters under three instructional conditions: physical match, name match, and rule match (in that order). A block of trials in the physical or name match conditions consisted of 72 pairs of letters; a block of trials in the rule match condition consisted of 64 pairs of letters. Subjects completed one block per condition per day. The stimuli themselves were various combinations of the letters A, E, H, and T. The order of pairs within a given block was randomized with respect to the occurrence of any letter or letter pair. In addition, the order of responses was randomized within each block with the restriction that the subject (if responding correctly) did not repeat the same response more than three consecutive times. Each subject repeated the procedures on the second day (with, of course, a different stimulus sequence).

Variables. The principal data were the mean RTs of correct responses for the four baseline measures which included: "same" judgments for the

physical, name, and category conditions and a "different" judgment calculated across conditions. In addition, two difference scores were calculated: a name minus physical match score and a category minus name match score. The percentage of errors and mean error time across conditions were also obtained.

Meyer task

Procedure. Each trial began with two central fixation points displayed on the CRT, one point above the other, which served as a cueing signal. After a foreperiod delay, the top point was replaced by the first string of capital letters; the subject judged whether or not it was a word. As soon as the subject responded, the top letter string was removed, and there was a short delay (500 msec) followed by the second string of letters. The subject again judged whether it was a word. Reaction time was measured separately for each string from the stimulus onset to the response. The intertrial interval was approximately two seconds. The following schematic shows the events in a typical trial:

Subject shown letter string	→	Subject responds	→	Subject shown letter string	→	Subject responds
GRINT	→	"nonword"	→	PRINT	→	"word"
FENCE	→	"word"	→	HENCE	→	"word"
MINT	→	"word"	→	PINT	→	"word"
GUIDE	→	"word"	→	LIED	→	"word"
BEARD	→	"word"	→	GONE	→	"word"

Stimuli and Design. During the eight test blocks (four on each day), each subject classified a total of 320 letter strings; each block, therefore, contained 40 strings or 20 trials of two strings each. Of the 160 pairs, 64 were word-word (WW) pairs, 32 were nonword-word (NW) pairs, 32 were WN pairs, and 32 were NN pairs. Among the 64 WW pairs, there were 16 each of the following types:

1. Graphemically and phonemically similar (e.g., FENCE-HENCE),
2. Graphemically similar, phonemically dissimilar (e.g., MINT-PINT),
3. Graphemically dissimilar, phonemically similar (e.g., GUIDE-LIED),
4. Graphemically and phonemically dissimilar (e.g., BEARD-GONE).

In addition to the WW pairs, there were a variety of pairs involving nonwords that followed the general rules of English orthography and phonology. There were four types of such pairs:

1. Word-nonword, graphemically similar (e.g., DRUNK-FRUNK),
2. Word-nonword, graphemically dissimilar (e.g., BOOT-ZAS),
3. Nonword-word, graphemically similar (e.g., GRINT-PRINT),
4. Nonword-word, graphemically dissimilar (e.g., MENTH-JOY).

It should be noted that graphemically similar nonwords could be pronounced like their word counterparts; thus, there was probably a degree of phonemic similarity in the first and third types above. Finally, there were two types of nonword-nonword pairs, those that looked similar (e.g., GRAT-TRAT), and those that looked dissimilar (e.g., DOPLY-TUCKEL). Each test block contained two examples of each of the 10 pair types; the order of pair types was randomized within each block with the restriction that the subject (if responding correctly) did not repeat the same response (i.e., word or nonword) more than three consecutive times nor was he exposed to more than three consecutive pairs with the same graphemic-phonemic relationship.

<u>Variables.</u> The principal data were the mean RTs of correct responses for the first letter string (word or nonword) of each pair and for the second letter string in each of the ten WW, WN, NW, and NN types. Several parameters were derived that involved the four WW pairs. A "phonemic facilitation" estimate was obtained by calculating the difference between the mean RT from the second member of the phonemically and graphemically similar pairs and the control (both dimensions dissimilar) condition; a "graphemic interference" score was calculated by finding the difference between the mean RT from the second member of the phonemically dissimilar, graphemically similar pairs and the control. Both of these parameters were reported by Meyer et al. and were

computed for the purpose of comparison. In addition, a third estimate was made to determine if phonemic similarity alone would facilitate recognition; this estimate was obtained by calculating the difference between mean RT from the second member of the phonemically similar, graphemically dissimilar pairs and the control. Percentage of errors and mean error time across conditions were also calculated.

Baron task

Procedure. Each trial began with a fixation point which served as a cueing signal. After a foreperiod delay, the subject was presented with a four-word phrase. The subject judged whether or not the phrase made sense (as a function of the instructional condition). Intertrial intervals were approximately three seconds. The following schematic shows the events in a typical trial:

	Subject shown phrase		Subject responds
SH	The sky is cloudy	→	"sense"
	It's knot so	→	"nonsense"
HN	See water is salty	→	"sense"
	The knife is pull	→	"nonsense"
SN	Please cash my check	→	"sense"
	A deck of carts	→	"nonsense"

Stimuli and Design. Each subject performed in three conditions, the conditions defined by the stimulus array and the instructional set:

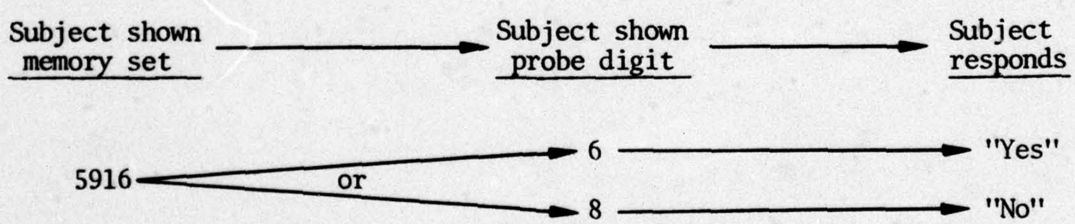
1. SH condition, where the stimulus phrases were either sense (S) or homophone (H) and subjects were told to call H phrases nonsense;
2. HN condition, where the phrases were either H or pure nonsense (N) and subjects were told to call H phrases sense; and
3. SN, where the phrases were either S or N and subjects were simply instructed to judge the phrases as sense or nonsense.

Each subject completed two blocks of 20 trials per block in each condition on each day. Within a block of trials, the order of phrases was randomized.

Variables. The basic data were the mean RTs for each phrase type (S, H, and N) as a function of condition (SN, SH, and HN). The data were combined to generate overall condition times (i.e., SN time, SH time, and HN time). Also, the ratio of SH time to HN time was calculated. This ratio was used to categorize subjects as either visual (low ratio) or phonemic (high ratio). Both percentage of errors and mean error time across conditions were also calculated.

Sternberg task

Procedure. Each trial began with the presentation of the positive memory set. On each trial, a set of digits, ranging randomly over trials from one to four different digits, was displayed for a duration directly proportional to the length of the set (one sec. per number). At the end of this interval, the displayed set was removed, and a cueing signal appeared, followed by the presentation of a single probe digit. The subject decided whether or not the probe digit was a member of the positive set. Reaction time was measured from onset of the probe digit to the response execution. The intertrial interval was approximately two seconds. Again, below are events in a typical trial:



Stimuli and Design. Each subject completed a total of 10 test blocks (five on each day), each block consisting of 25 trials. The 25 trials in each block were composed of cases for memory set sizes of one to four items. Half of the trials were positive (i.e., the probe item was contained in the memory set) and half were negative. Individual trials were generated so that, for the positive items, there was no bias with respect to serial position in the memory set of the probe digit. The sequence of trials with-

in each block was randomized with respect to memory set size, with the restrictions that the subject was not exposed to more than three consecutive trials of a particular memory set size, nor was he required (if responding correctly) to repeat the same response more than three times in succession.

Variables. The principal data were the mean RTs of correct responses for each memory set size. The RTs for both positive and negative responses were used to calculate (for each subject) the slope and intercept of the best-fitting linear function relating mean RT to memory set size. In addition, percentage of errors and mean error time across conditions were also computed.

Juola task

Procedure. The procedures for the two conditions (word and category) were identical to each other and essentially equivalent to the procedures used in the Sternberg task described above. Each trial began with the presentation of the positive memory set. On each trial, a set of words, ranging randomly over trials from one to four different words, was displayed for a duration directly proportional to the length of the list. At the end of this interval, a cueing signal appeared, followed shortly by the presentation of a single probe word. The subject decided whether or not the probe word was included in the memory set (for word condition) or was an exemplar of one of the categories in the memory set (for the category condition). Reaction time was measured from the onset of the probe word to the response execution. The intertrial interval was approximately two seconds. Below are events in a typical trial:

	Subject shown memory set	→	Subject shown probe word	→	Subject responds
Word	{clothing insect	→ or	insect	→	"Yes"
		→	vehicle	→	"No"
Category	{clothing color	→ or	pink	→	"Yes"
		→	car	→	"No"

Stimuli and Design. Each subject completed a total of eight test blocks (four on each day), four blocks per condition (the word condition blocks were presented first on each day), each block consisting of 30 trials. The 30 trials in each block were composed of cases for memory set sizes of one to four words. Half of the trials were positive and half were negative. Individual trials were generated so that there was no positional bias in the presentation of positive members of the memory set. The sequence of trials was randomized with respect to memory set size, with the restrictions that the subject was not exposed to more than three consecutive trials of a particular memory set size, nor was he required to repeat the same response (if responding correctly) more than three consecutive times.

The words themselves consisted of nine category names (from which the memory sets in both conditions were composed and from which the probe stimuli in the word condition were chosen) and nine exemplars of each category as shown below:

A. Color	(1) Blue, (2) Red, (3) Green, (4) Yellow, (5) Black, (6) White, (7) Pink, (8) Brown, and (9) Gray
B. Bird	(1) Robin, (2) Sparrow, (3) Eagle, (4) Crow, (5) Duck, (6) Hawk, (7) Parrot, (8) Dove, and (9) Owl
C. Tree	(1) Oak, (2) Maple, (3) Cedar, (4) Elm, (5) Pine, (6) Spruce, (7) Birch, (8) Poplar, and (9) Fir
D. Fruit	(1) Lemon, (2) Plum, (3) Apple, (4) Peach, (5) Cherry, (6) Grape, (7) Pear, (8) Banana, and (9) Lime
E. Fish	(1) Shark, (2) Trout, (3) Cod, (4) Salmon, (5) Sardine, (6) Perch, (7) Tuna, (8) Whale, and (9) Flounder
F. Insect	(1) Fly, (2) Ant, (3) Bee, (4) Spider, (5) Moth, (6) Flea, (7) Termite, (8) Beetle, and (9) Wasp
G. Clothing	(1) Shirt, (2) Pants, (3) Shoes, (4) Blouse, (5) Coat, (6) Dress, (7) Hat, (8) Jacket, and (9) Gloves
H. Family	(1) Aunt, (2) Uncle, (3) Father, (4) Mother, (5) Brother, (6) Sister, (7) Cousin, (8) Niece, and (9) Nephew
I. Vehicle	(1) Car, (2) Bus, (3) Plane, (4) Truck, (5) Bike, (6) Train, (7) Wagon, (8) Taxi, and (9) Boat

Variables. The principal data were the mean RTs of correct responses for each memory set size in each condition. For each subject, the slope and intercept of the best fitting linear function were calculated for both positive and negative responses. Again, the error parameters were computed.

Clark and Chase task

Procedure. Each trial began with a central fixation point which served as a cueing signal. After a foreperiod delay, the subject was presented simultaneously with a sentence on the left and a picture on the right side of the display. The subject read the sentence and decided whether or not the sentence was an accurate description of the picture. Reaction time was measured from the stimulus onset to response; the intertrial interval was approximately two seconds. Below are events in a typical trial:

Subject shown sentence		Subject responds
Star isn't below cross $\overset{+}{*}$	$\longrightarrow$	"False"
Cross isn't below star $\overset{+}{*}$	$\longrightarrow$	"True"

Stimuli and Design. Each subject completed 10 blocks of 16 trials (five blocks per day), each block consisting of a different random order of 16 displays. The 16 displays consisted of eight sentences paired with one of two pictures.

The sentences included "Star is above plus "; "Star is below plus "; "Star isn't above plus "; "Star isn't below plus ", as well as the four corollary sentences with "star" and "plus" interchanged. The pictures were either an asterisk (star) directly above a plus, $\overset{*}{+}$, or a plus above an asterisk, $\overset{+}{*}$. Each of the 16 displays could be characterized along three dimensions:

1. whether it was a positive (P) or a negative (N) sentence (e.g., "Star is (isn't) above plus.");
2. whether the sentence was a true description of the picture (T) or a false description (F); and

3. whether the sentence contained the word "above" (A) or the word "below" (B).

Thus, in each block of 16 displays, there were two examples of each of eight display types: PTA, PTB, PFA, PFB, NTA, NTB, NFA, and NFB.

__Variables.__ The primary data were the mean RTs of correct responses to each of the eight display types. These means were used to calculate estimated values for each of Clark and Chase's four parameters previously described. These parameters were estimated according to the following equations:

$$
\begin{array}{lllllllll}
\underline{a} = & \frac{1}{4}\,[(\mathrm{PTB}-\mathrm{PTA}) & + & (\mathrm{PFB}-\mathrm{PFA}) & + & (\mathrm{NTB}-\mathrm{NTA}) & + & (\mathrm{NFB}-\mathrm{NFA})] \\
\underline{b}\ \&\ \underline{d} = & \frac{1}{4}\,[(\mathrm{NTA}-\mathrm{PFA}) & + & (\mathrm{NTB}-\mathrm{PFB}) & + & (\mathrm{NFA}-\mathrm{PTA}) & + & (\mathrm{NFB}-\mathrm{PTB})] \\
\underline{c} = & \frac{1}{4}\,[(\mathrm{PFA}-\mathrm{PTA}) & + & (\mathrm{PFB}-\mathrm{PTB}) & + & (\mathrm{NTA}-\mathrm{NFA}) & + & (\mathrm{NTB}-\mathrm{NFB})] \\
\underline{t}_o = & \frac{1}{2}\,[\ \mathrm{PFA} & + & \mathrm{PTB} & - & \mathrm{NTB} & + & \mathrm{NFA}\ ]
\end{array}
$$

Both the percentage of errors and mean error time across conditions were also obtained.

__Collins and Quillian task__

__Procedure.__ Each trial began with a left fixation point which served as a cueing signal. After each fixed delay, the point was replaced by a four- or five-word sentence. The subject judged whether the sentence was generally true or false. The sentence remained visible until the subject made a response. This cycle was repeated with an intertrial interval of approximately two seconds. Reaction time was measured as the interval between stimulus onset and response initiation. Below are events in a typical trial:

__Subject shown sentence__ ⟶ __Subject responds__

S2 { A canary is an animal ⟶ "True"

P2 { A canary has skin ⟶ "True"

F1 { A canary is a fish ⟶ "False"

Stimuli and Design. During the four test blocks (two on each day), each subject verified a total of 144 sentences; each block contained 36 sentences of which half were true and half were false. Each block contained six exemplars of each of six sentence types. These sentence types stated either property (P) relations (e.g., "Roses are red ") or superset (S) relations (e.g., "Roses are flowers "). Each P or S sentence had a number added to it; this indicated the number of levels necessary to "move through" to decide whether the sentence was true. For example, "A canary can sing " was a P0 sentence; "A canary can fly " was a P1 sentence (since it was presumed that flying is a property of birds); and "A canary has skin " was a P2 sentence (since it was presumed that "has skin" is a property of animals). Similarly, "A canary is a canary " was an S0 sentence; "A canary is a bird " was an S1 sentence; and "A canary is an animal " was an S2 sentence.

The sentences were generated from 12 separate "information structures". Each of these structures consisted of a three-level hierarchy with the words ordered by a subset relationship and properties attributable to particular subset levels. For example, the sentences from one structure were as follows:

S0: A trout is a trout.
S1: A cod is a fish.
S3: A salmon is an animal.
P0: A shark is dangerous.
P1: A flounder can swim.
P2: A barracuda breathes.

These same 12 structures were also used to generate negative sentences by rearranging property and superset information among structures. Each block of 36 trials used the positive and negative sentences derived from three of the structures.

Variables. The principal data were the mean correct verification times for each of the six positive sentence types and a mean correct verification time for negative sentences across levels. The best-fitting linear function was calculated to fit the relationship between levels of sentences and mean

RTs for the positive S and P sentences. Both the percentage of errors and mean error time across conditions were computed.

Shepard and Teghtsoonian task

Procedure. In this self-paced (within limits) task, subjects decided whether or not they remembered having seen each number earlier in the series. Subjects were allowed to proceed through the list at their own speed (maximum of 10 seconds), but were not permitted to take notes. No feedback as to correctness of responses was provided. Below are events in a typical trial:

Subject shown number		Subject responds
221	⟶	"new"
⋮		⋮
221	⟶	"old"

Stimuli and Design. Subjects were exposed to one list per day; the lists were each 101 items in length. The three-digit numbers included in each list were randomly selected from the total population from 100 through 999 (barring triples). With a single exception, every number in a given list was presented exactly twice. The second occurrences of the numbers were placed so that several lags between presentation and test were represented. The lags used were 1, 2, 4, 8, 12, 16, 20, 24, 30, and 36 items, with five exemplars of each lag in a given list. The probability of a "new" item was slightly higher than .5 in the first part of the list, and the probability of an "old" item was slightly higher than .5 near the end of a list.

Variables. Several measures were derived from this task. These measures fell into two categories. First, there were "traditional" retention parameters, and second, there were parameters derived from signal detection theory. Of the traditional measures, the two employed here were proportion correct (i.e., X/101) and a two-parameter estimate of the best-fitting curve

for the probability correct by lag function. This function was characterized by the least squares estimates for A and B in the equation:

$$y = Ax^B, \text{ where } A = \text{intercept and } B = \text{exponent.}$$

These parameters were calculated for each subject for each day.

The signal detection parameters were derived from two observed scores, namely:

(a) probability that the subject responded "old" when the stimulus was old (P "O" $|$ O) or "hits"

(b) probability that the subject responded "old" when the stimulus was new (P "O" $|$ N) or "false alarms".

The signal-detection discrimination parameter d' was calculated as the normal deviate of (a) plus the normal deviate of (b). Beta was calculated as the normal ordinate of (1 - a) divided by the normal ordinate of (b).

RESULTS AND DISCUSSION

The results of three sets of analyses are presented below. The first set deals with the replicability of previous experimental work with highly similar paradigms. The principal issue is whether or not the major findings were supported in terms of group main effects, despite changes in the specific implementation of each task. The second set of analyses is concerned with the information-processing parameters obtained from the tasks. Various aspects of the measures are examined, including their reliability, practice effects, descriptive statistics, and the character of the response distributions in the subject population. The final set of analyses deals with validity-type issues. The inter- and intra-task correlation matrices are presented and discussed in terms of the constructs represented.

Replications of Group Effects

Posner task. Figure 1 presents the results for the Posner task in the form of a tree diagram. These results indicate that the pattern of the Posner and Mitchell (1967) data was replicated in the current experiment. The data, particularly for Day 2 performance, are quite similar to the Posner and Mitchell results in spite of differences in the size and appearance of the letter displays used in the two experiments. The discrepancy in mean reaction time (RT) to respond "different" may be attributable to differences in the approach used to calculate the measure. In the Posner and Mitchell study "different" RT was based upon judgments from the "category match" condition only, whereas in the current work the measure is based upon pooled judgments across the three conditions.

In general, the data from both experiments show a monotonically increasing RT as a function of depth of abstraction. With the exception of one subject, all of the node-to-node processing times are monotonically increasing in the current study. Day 2 performance is consistently faster than Day 1 performance at all levels of abstraction by about 60 msec.

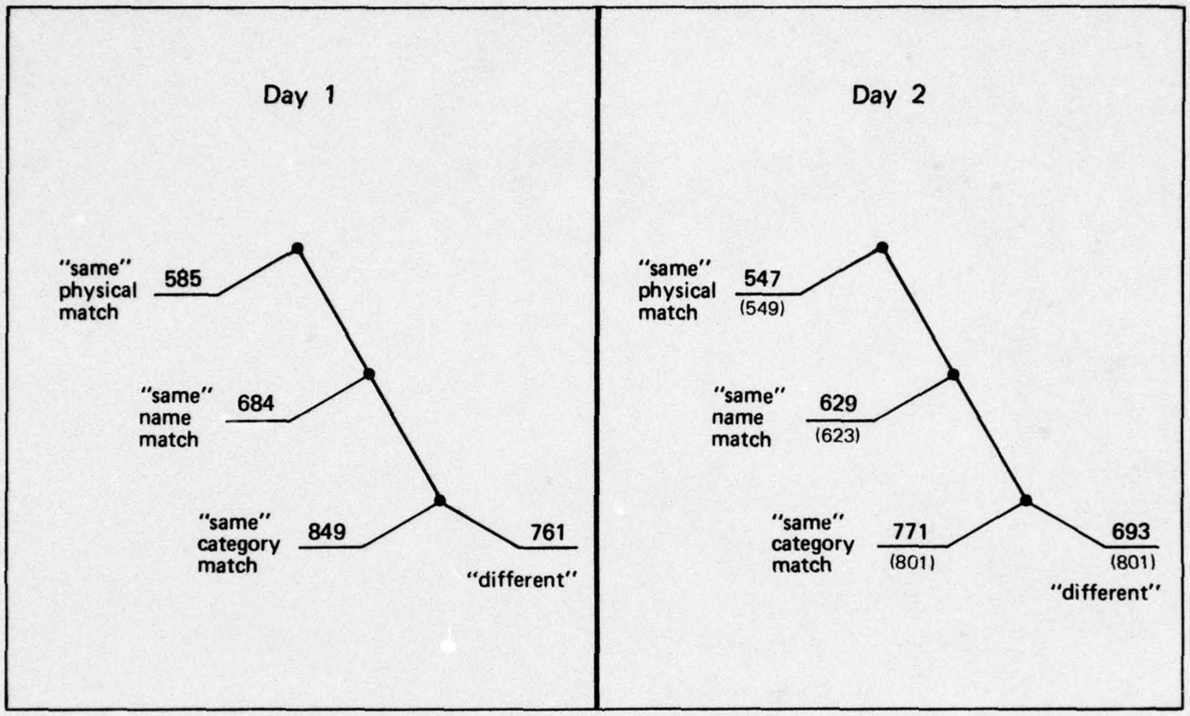


Figure 1. Tree diagram for letter classification task*

* Tree diagrams for Day 1 and Day 2 data. Numbers refer to mean RTs (msec) for all subjects. Numbers in parentheses show response times obtained by Posner and Mitchell (1967).

Meyer task. Table 2 summarizes the mean RTs obtained for the Meyer task. As in Meyer et al. (1974), words are usually classified more quickly than nonwords. This is the case whether the item is the first or second letter string of a stimulus pair and is characteristic of both Day 1 and Day 2 performance.

Meyer et al. found that graphemic similarity together with phonemic similarity facilitated recognition (i.e., the difference between Type 4 and Type 1 stimulus pairs is positive), but that graphemic similarity alone inhibited performance (i.e., the difference between Type 4 and Type 3 stimulus pairs is negative). This pattern of results is replicated in Day 1 performance of the current experiment, the magnitude of differences being even larger than those reported by Meyer et al. Day 2 performance also shows a large facilitation effect when both phonemic and graphemic similarity are present, but the inhibition effect for graphemic similarity alone is essentially zero.

These results suggest that, on Day 2, subjects may have adopted a different strategy for responding to the second letter string of each pair. According to this strategy, subjects would decide, first, if the second letter string was both graphemically and phonemically similar to the first string. Second, if the strings were not similar, the subjects would determine if the second letter string was a word or a nonword. This strategy would account for the Day 2 results in which RTs to words that looked and sounded alike were fastest, followed by the other word conditions regardless of graphemic or phonemic similarity, and lastly all of the nonword conditions.

In the current experiment, the Meyer paradigm was modified to include a category of phonemically similar but graphemically dissimilar word pairs. It was expected that the presence of phonemic similarity alone would facilitate recognition. The results appeared to be counterintuitive; as can be seen in Table 2, phonemic similarity alone inhibited performance (i.e., the difference between Type 4 and Type 2 stimulus pairs was positive) in

Table 2

Parameter Estimates for Meyer Task

	Type of Stimulus Pair	Phonemic Relation	Graphemic Relation	Mean RTs (msec) Second Letter String: Meyer, et al. (1974)	Present Study: Day 1	Present Study: Day 2
(1)	Word-Word	Similar	Similar	589	628	586
(2)	Word-Word***	Similar	Dissimilar	-	688	659
(3)	Word-Word	Dissimilar	Similar	633	700	643
(4)	Word-Word	Dissimilar	Dissimilar	602**	664	642
(5)	Word-Nonword	Similar*	Similar	701	821	725
(6)	Word-Nonword	Dissimilar	Dissimilar	716	881	724
(7)	Nonword-Word	Similar*	Similar	601	686	615
(8)	Nonword-Word	Dissimilar	Dissimilar	565	713	620
(9)	Nonword-Nonword	Similar*	Similar	779	858	728
(10)	Nonword-Nonword	Dissimilar	Dissimilar	795	905	786

Differences

	Meyer, et al. (1974)	Day 1	Day 2
Type 4 - Type 1	13	36	56
Type 3 - Type 4	31	36	1

	Day 1	Day 2
Mean RT (msec) First Word	736	647
Mean RT (msec) First Nonword	916	756
Percent Errors	7.4	7.5
Mean RT Errors	875	666

* Nonwords may have been either phonemically similar or dissimilar to their mates.
** Mean of Meyer's Type 2 and Type 4.
*** This stimulus-type absent in Meyer's study.

both testing sessions, the magnitude of the difference being about the same on both days.

Baron task. The modifications of the Baron (1973; Baron & McKillop, 1975) procedure adopted for the present study preclude a direct comparison of obtained results. Specifically, the Baron (1973) procedure involved presenting all three phrase types (Sense, Nonsense, and Homophone) in all conditions; his conditions differed as a function of instruction (i.e., the first condition required subjects to judge sense or nonsense on the basis of appearance, and the second condition required subjects to make their judgments on the basis of an acoustic representation). In addition, subjects were administered several hundred trials per condition. The Baron & McKillop (1975) procedure used the 3 conditions employed in our study; however, in their study stimuli were presented in a list, subjects were required to mark each phrase, and the dependent measure was time to complete each list. Nevertheless, certain aspects of the data are interesting to compare with the results of our study (see Table 3).

Baron (1975) tested two primary hypotheses that have parallels in the present experiment. First, he argued that there would be no difference between H and N phrases in the "appearance" condition. In our terminology, this comparison can be translated into an examination of N phrases in the SN condition, and the H phrases in the SH condition. As can be seen in the table, we likewise found no differences (on either day of testing). His second major hypothesis was that the H phrases would take substantially longer than the S phrases in the "sound" condition. This translates into a comparison between the H phrases in the HN condition and the S phrases in the SH condition. Again, our results confirmed Baron's hypothesis.

Data from Baron & McKillop (1975) are also shown in Table 3. Unfortunately, they reported data only for the five lowest and five highest SH/HN ratio subjects. The times-per-item shown in Table 3 were calculated from their figures. These data are not directly comparable to those in

Table 3

Summary of Selected Measures for Baron Task in Mean RT/item (msec)

	Session 1		Session 2		
	Baron, et al. (1973)	Present study	Baron, et al. (1973)	Present study	Baron, et al. (1975)
S	1046	1126	981	1120	-
N	1106	1300	993	1277	-
SN	-	1205	-	1193	1508
S	875	1265	738	1149	-
H	1116	1302	1008	1221	-
SH	-	1289	-	1187	1663
H	935	1512	783	1348	-
N	1208	1665	997	1522	-
HN	-	1579	-	1423	1779

the current study for several reasons (e.g., differences in response made, sampling of "extreme" groups of subjects, etc.); however, it is interesting to note that the HN times are substantially greater than the SN and SH times in both studies. Also, it seems that with practice, our subjects make the SN decision (which is ambiguous in the sense that it could be made on the basis of the appearance or the sound of the phrases) at the same rate as the SH decision (which requires a "visual" judgment). This confirms Baron's hypothesis that the visual strategy is faster and more efficient than the phonemic strategy and probably uses information that is available "earlier" in phrase processing.

Sternberg task. The linear functions relating RT to memory set size for the Sternberg task are shown in Figure 2. The data for Day 1 and Day 2 performance are plotted separately. In both cases, RT increases linearly with set size, at the same rate for both positive and negative responses. The rate of increase is about 62 msec for each item in the positive set on Day 1; the intercept is about 490 msec. The function for Day 2 shows a general improvement in performance, with the slope about 48 msec per item and the intercept about 440 msec. These results support a serial exhaustive model of memory scanning and compare favorably with results obtained by Sternberg (1975). In his experiments, the slope of the RT function was about 38 msec per item and the intercept was about 400 msec. The subjects in Sternberg's study were tested over a large number of trials so that the flatter slope and lower intercept values obtained in his study may be attributed, at least in part, to more practice.

Juola task. Figure 3 presents the linear functions relating RT to memory set size for word and category conditions. Separate functions were calculated for positive and negative responses and for Day 1 and Day 2 sessions. In both conditions, RT increases linearly with the number of memory set items; this is true for both positive and negative responses. The intercepts for word and category conditions are similar, but the slopes for categories are considerably steeper than those for words. The slope

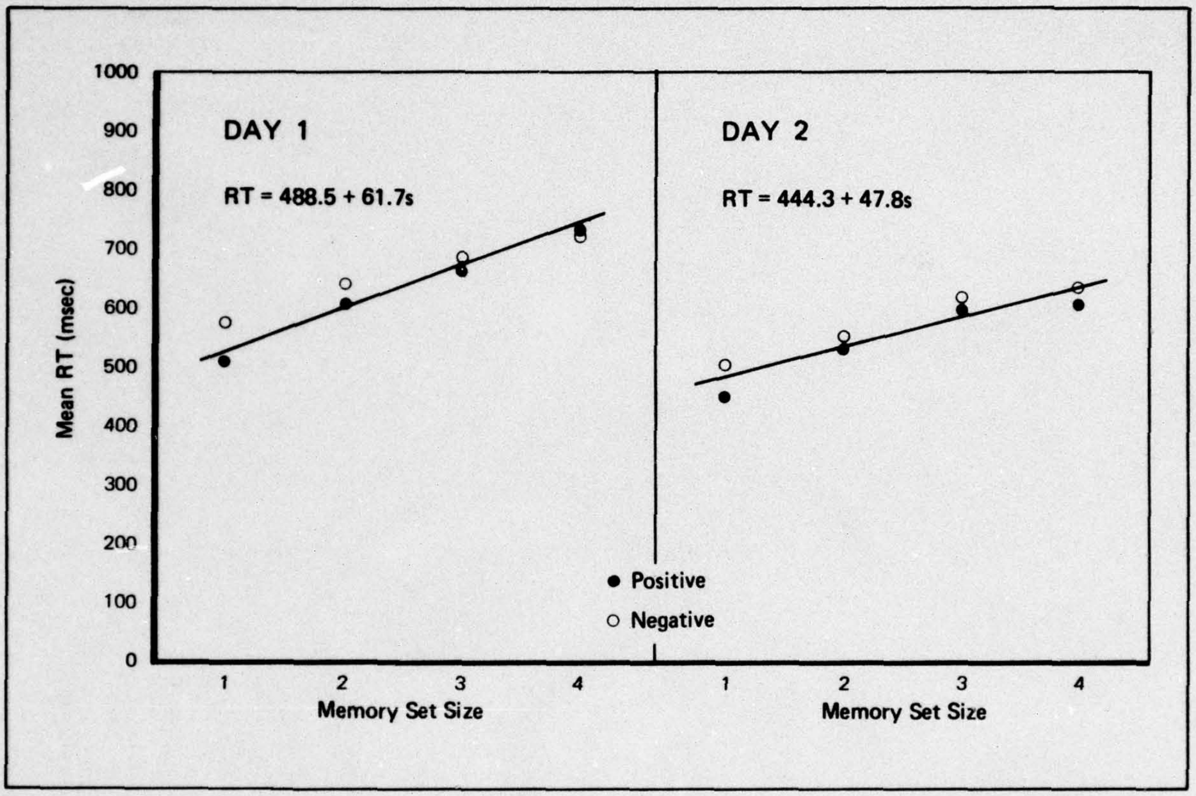


Figure 2. Day 1 and Day 2 mean RTs on positive and negative trials for Sternberg task.

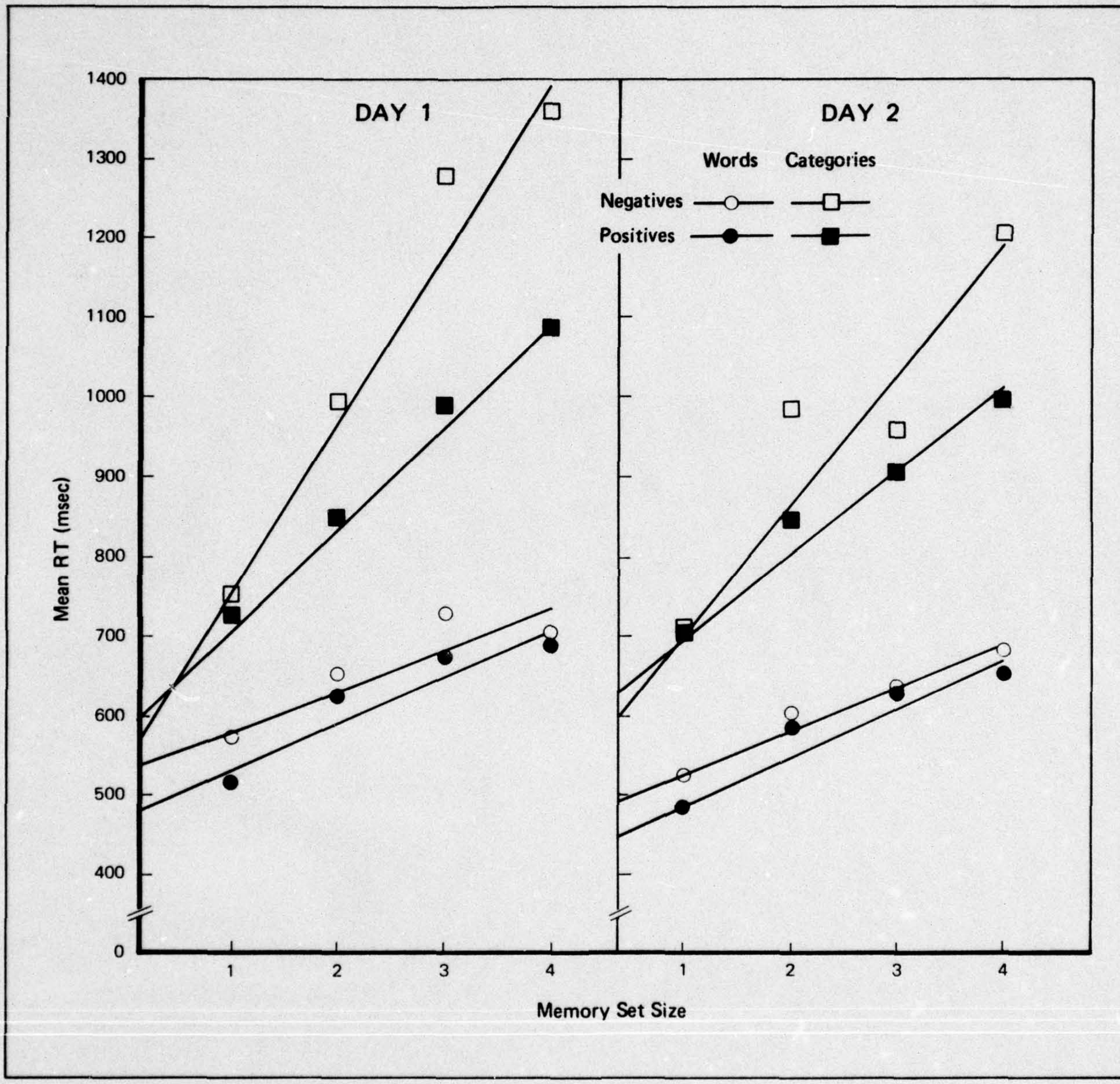


Figure 3. Best-fit regression lines for Juola and Atkinson task.

and intercept parameters for the current study and those obtained by Juola and Atkinson (1971) are shown in Table 4. The pattern of their results was replicated in the current experiment; the parameter estimates for both testing sessions are quite similar to theirs despite modifications in the construction of stimulus items in the current study.

An analysis of the slope parameters provides some insight into the nature of the search processes in the two conditions. The slopes for positive and negative responses in the word condition are nearly equal indicating that the scanning process for words was exhaustive: all possible pairwise comparisons were made before deciding whether or not a match had occurred. In the category condition, however, the slope for positive responses is considerably flatter than that for negative responses. This would suggest a scanning process for categories that self-terminated when an item in the memory set was found to match the target. Although performance in both conditions improved from one day to the next, these differences in scanning were apparent in both sets of results.

Another aspect of the memory search task is the comparison operation itself. An analysis of the slope parameters also provides some insight into this operation in the two conditions. One alternative was that in the category condition, subjects would convert the target word to its appropriate category name and then compare this name with all of the items in the memory set. For example, if the memory set was "COLOR CLOTHING BIRD" and the target word "ROBIN", the subject would convert the target word to a category ("a robin is a bird") and then search the memory set for "BIRD". If the search process operated in this manner, the slope of the RT function for categories would be similar to that for words but with a higher intercept. On the other hand, subjects might decide, for each member of the memory set, whether the target word was a member of that catagory (e.g., Is a robin a color? Is a robin a clothing? Is a robin a bird?). In this case, the slope for categories would be greater than that for words but the two functions would have similar intercepts. The results in Figure 3 support the latter of these alternatives; the slopes of the category functions for

Table 4

Slopes and Intercepts (msec) of the Best-Fitting Linear Functions Relating Mean RTs to Memory Set Size in Juola Tasks

Measure	Present Study Day 1	Present Study Day 2	Juola (1971) Data
Word task			
Slopes			
Positive trials	57	56	49
Negative trials	46	51	26
Mean	52	54	38
Intercepts			
Positive trials	481	446	543
Negative trials	545	485	617
Mean	513	466	580
Categorization task			
Slopes			
Positive trials	130	92	89
Negative trials	209	147	111
Mean	170	120	100
Intercepts			
Positive trials	605	634	670
Negative trials	578	586	653
Mean	592	610	662

both testing sessions are, on the average, two to three times greater than the slopes of the word functions.

Clark task. Table 5 presents the mean RTs by sentence type for Day 1 and Day 2 and the observed latencies obtained by Clark and Chase (1972, Experiment 1). Day 1 and Day 2 performance is shown in Figure 4 as a function of whether the sentence was an affirmative or negative sentence, whether it contained "above" (a) or "below" (b), and whether it was true or false with respect to the picture. These results compare favorably with those of Chase and Clark in terms of both overall RTs and pattern of the data.

Parameters for predicting verification RTs for each testing session were calculated according to the method described previously. The estimates for Day 1 performance were: "below" time 136 msec, "negation" time 829 msec, "comparison" time 200 msec, and "base" time 1735 msec. The estimates for Day 2 were 110 msec, 685 msec, 146 msec, and 1489 msec, respectively. The first three parameters were found to account for 98+ percent of the variance among the eight means for each session. These results replicated those of Clark and Chase who obtained parameter estimates of 93 msec, 685 msec, 187 msec, and 1763 msec, respectively, accounting for more than 99 percent of the variance among means.

In order to further test the validity of the four parameter model, subjects were divided into odd- and even- numbered groups. The four parameters based on the observed RTs of one group were used to predict RTs for the eight sentence types of the other group. The correlations between observed and predicted latencies for the two groups were both greater than $r = .97$, indicating that the model was quite powerful in terms of accounting group performance.

Collins and Quillian task. The functions relating RT to sentence level for Day 1 and Day 2 are shown in Figure 5. Although our subjects were generally faster at confirming sentences than those used by Collins and Quillian (1969), the patterns of results in the two studies were quite similar. RT

Table 5

Breakdown of Latencies for Eight Types of Sentences from Clark and Chase Task.

Sentence Type			Sentence	Latency Components	Observed RT (msec) Present Study Day 1	Day 2	Clark and Chase (1972) observed data
Positive	True	above	A is above B	t_0	1689	1432	1744
		below	B is below A	$t_0 + a$	1787	1556	1875
	False	above	B is above A	$t_0 + c$	1896	1597	1959
		below	A is below B	$t_0 + a + c$	2143	1754	2035
Negative	True	above	B isn't above A	$t_0 + b + c + d$	2778	2312	2624
		below	A isn't below B	$t_0 + a + b + c + d$	2755	2321	2739
	False	above	A isn't above B	$t_0 + b + d$	2538	2146	2470
		below	B isn't below A	$t_0 + a + b + d$	2759	2297	2520

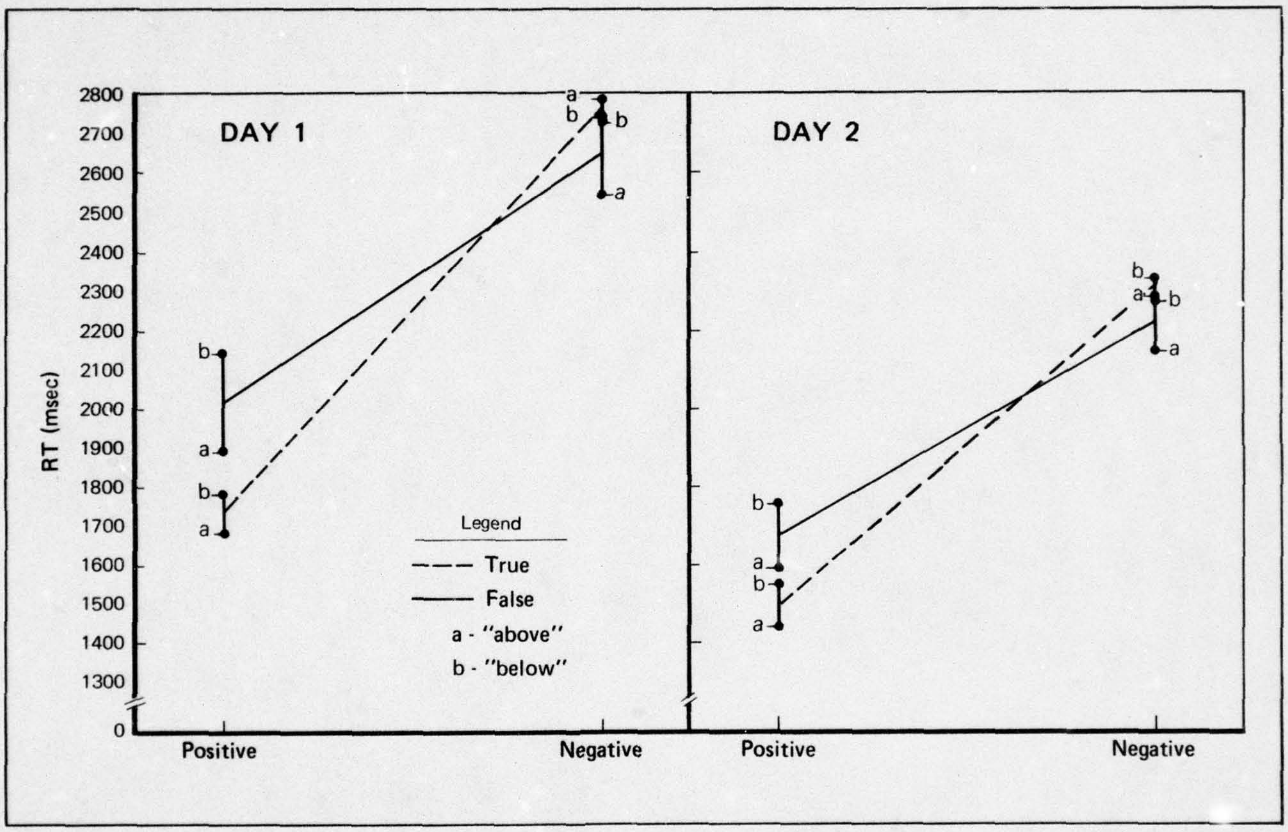


Figure 4. Mean RTs (msec) for Clark and Chase Task.

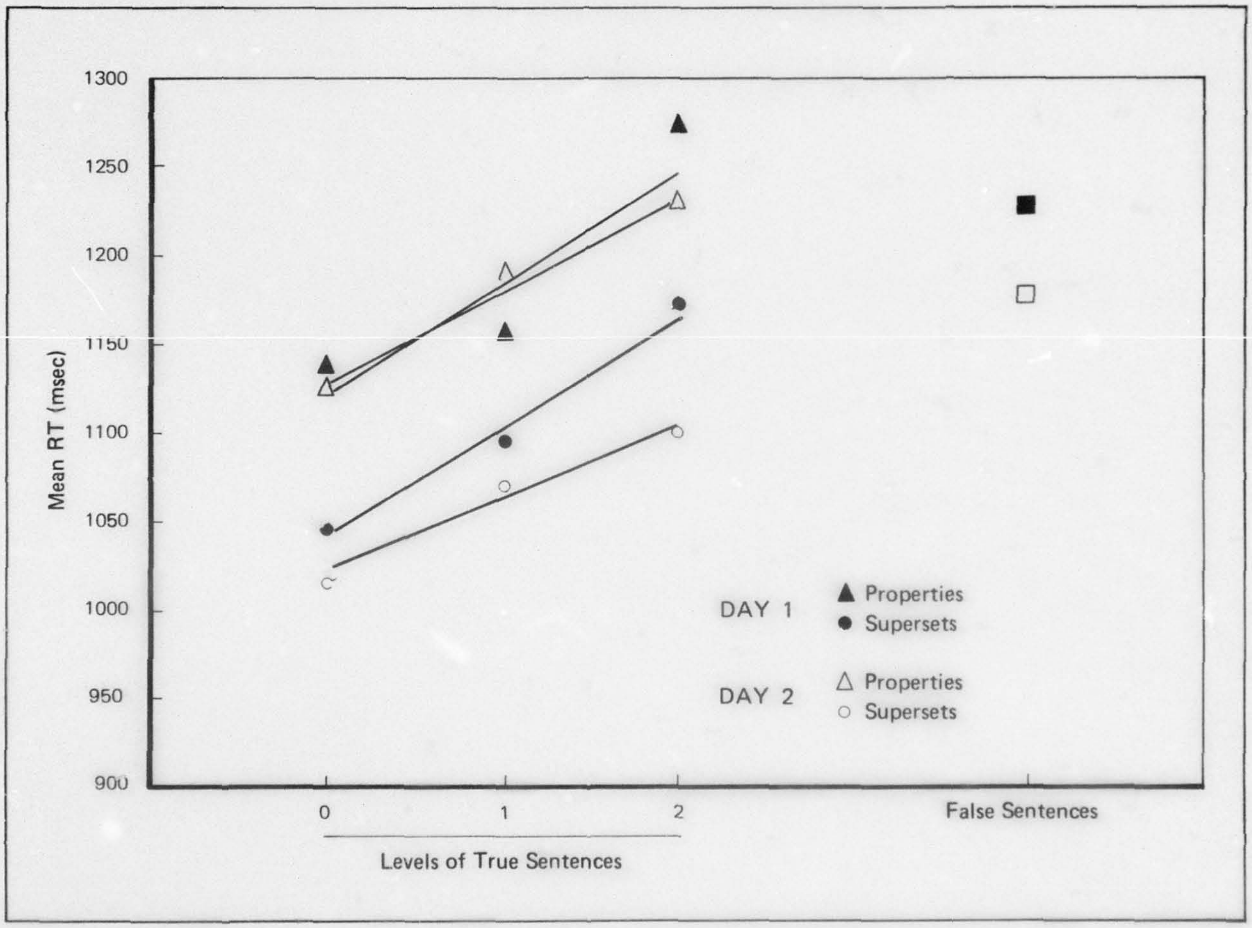


Figure 5. Mean RTs (msec) and best-fit regression lines for Collins and Quillian task.

increased linearly with the number of levels separating memory nodes when retrieving either superset or property information. For Day 1 performance, the slope of the two functions is about 66 msec, which represents an estimate of the time needed to retrieve superset information from LTM. For Day 2, this figure is about 48 msec. If the RT functions for property and superset are assumed to be parallel, the difference in intercepts serves as an estimate of the time to retrieve property information; that is, about 125 msec on Day 1 and 80 msec on Day 2. Our obtained slope and intercept estimates are generally faster than those obtained by Collins and Quillian who reported a slope of 75 msec and a difference in intercept of 225 msec.

Despite alternative explanations that could account for these data, the present study replicated the finding that, for these stimuli, property and superset functions are parallel, with different intercepts. Whether semantic memory effects are explained in terms of a hierarchical structure or some other model, the process of moving from one memory node to another was found to take a predictable amount of time. For present purposes, this RT is being used as a measure of scanning through LTM.

Shepard and Teghtsoonian task. Figure 6 summarizes the effect of delay on the accuracy of classifying an old three-digit number as "old". Separate functions are presented for Day 1 and Day 2 and for the two sessions combined. Although the curves are very rough, they represent the course of forgetting as delays get longer. The likelihood of recognizing an old number as "old" is almost perfect when the number was just seen. The probability drops to about .8 with delays of 8 intervening items, then to about .7 with delays of 20 to 36 items. Even with the maximum delay of 36 items, the probability of correctly recognizing an old number was well above chance level. These results (Table 6) parallel those of Shepard and Teghtsoonian (1961) despite changes in procedure in the current experiment. They reported "hit" probabilities (of recognizing an old number as "old") of about .77 with delays of 6 or 7 items and about .65 for delays of about 35 items. The probability of a "false alarm" (i.e., classifying a new number as "old") in

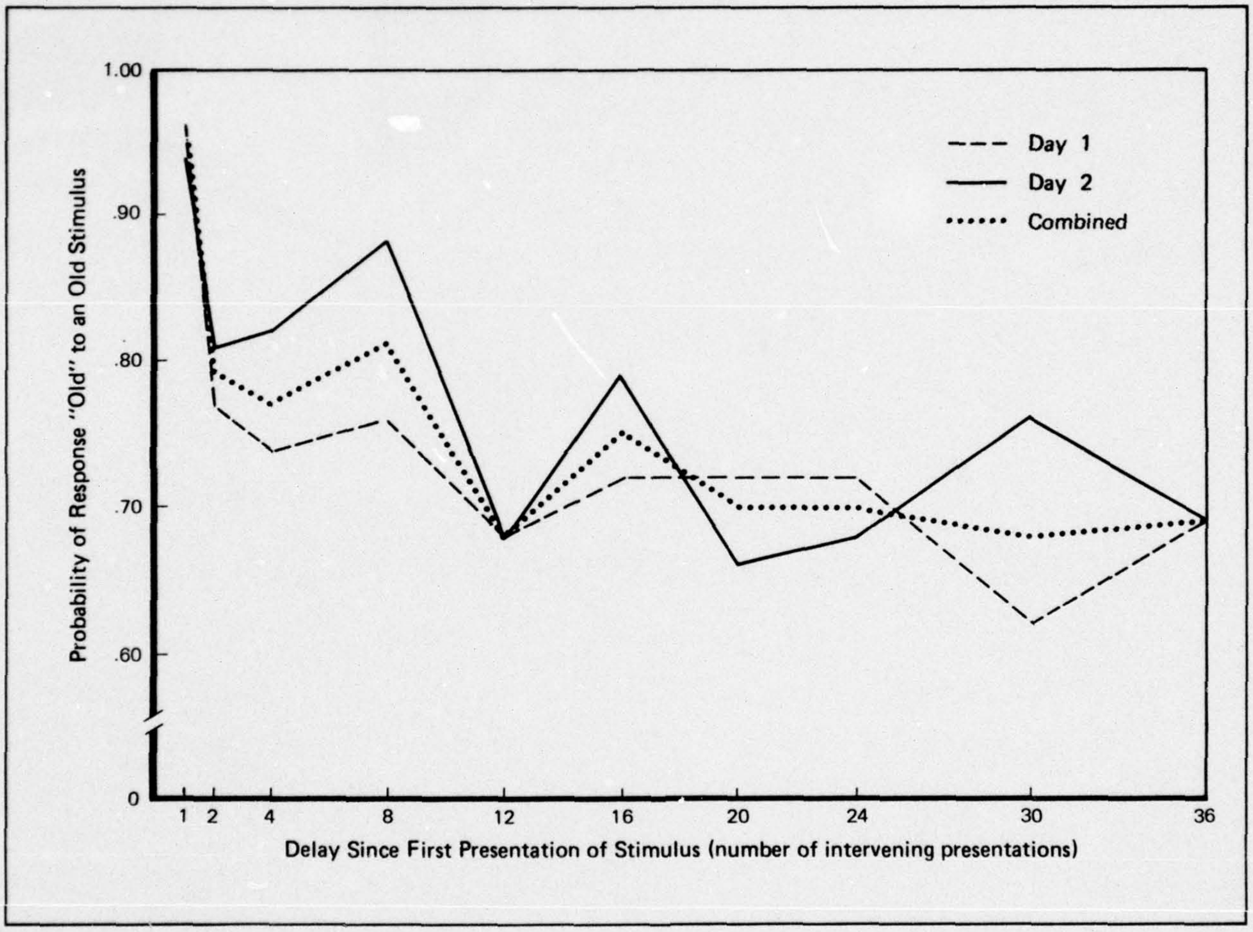


Figure 6. Lag function for Shepard and Teghtsoonian task.

Table 6

Selected Shepard and Teghtsoonian Task Parameters

Variables	Present study Day 1	Present study Day 2	Shepard et al. (1961)
Hits (p("O" given O))	.73	.77	.77*
False Alarms (p("O" given N)	.28	.31	.23
d′ (in Z scores)	1.284	1.335	--
Beta (Z score ratio)	1.084	0.929	--

* Based on lags of six and seven items.

their experiment averaged about .23. This figure compares favorably with "false alarm" rates of .28 and .31 for the two sessions in the current experiment.

Individual Measures

Overview. The above results indicate that, for the most part, the major group effects were replicated in each paradigm. Thus, there is demonstrated empirical and theoretical support for the information processing constructs contained in the tasks. However, the value of the paradigms for an assessment battery depends primarily on the measures derived from them and the properties of these measures when considered as potential individual difference variables. This distinction between task effects and measurement properties is particularly important in the present context since most of the paradigms were not originally generated for the study of individual differences; the scientists were primarily concerned with uncovering different aspects of the human information processing system. Similarly, these paradigms have not previously been considered as tests per se; no thought has been given to typical test development issues. Finally, the distinction between group effects and individual measures is critical in that several theoretically independent measures can be obtained from each task. For example, the Shepard and Teghtsoonian task results can be described by a number of different parameters: the "standard" measure of proportion of correct items (or, more finely, proportion of "hits" and "false alarms"), the two parameters of the exponential equation that is the best fit to the probability-correct-by-lag function, and the signal-detection-theory parameters d' and β.

Given these considerations, a set of 40 variables was selected for detailed examination. These variables are shown in Table 7. Also shown in this table are the theoretical operations that these variables are hypothesized to measure. The operations were derived primarily from Table 1 above, which described the operations involved in each task condition. As can be seen in Table 7, there are several "redundancies" in the operations

Table 7
Operations for Each Task Measure

Measure	Operations
POSNER PHYSICAL	• Encode letters • Compare representations of letter patterns • *Select and execute response*
POSNER NAME	• Encode letters • Retrieve name from LTM • Compare representations of letter names • Select and execute response
POSNER CATEGORY	• Encode letters • Retrieve name from LTM • Retrieve category from LTM • Compare representations of letter categories
POSNER NAME MINUS PHYSICAL	• Retrieve name from LTM
POSNER RULE MINUS NAME	• Retrieve category from LTM
MEYER WORD	• Encode letter string • Construct phonemic or graphemic representation • Search LTM for "word" • Select and execute response
MEYER NONWORD	• Encode letter string • Construct phonemic or graphemic representation • Search LTM for "word" • Search LTM for "word" rules • Select and execute response
BARON SH	• Encode semantic phrases • Construct visual representation • Search LTM for "phrase" rules • Select and execute response
BARON HN	• Encode semantic phrases • Construct acoustic representation • Search LTM for "phrase" rules • Select and execute response
BARON SN	• Encode semantic phrases • Construct visual or acoustic representation • Search LTM for "phrase" rules • Select and execute response
STERNBERG INTERCEPT POSITIVE, NEGATIVE RESPONSES	• Encode target number • Select and execute response
STERNBERG SLOPE POSITIVE, NEGATIVE RESPONSES	• Compare numbers (serial)
JUOLA WORD INTERCEPT POSITIVE, NEGATIVE RESPONSES	• Encode target word • Select and execute response
JUOLA WORD SLOPE POSITIVE, NEGATIVE RESPONSES	• Compare words (serial)
JUOLA CATEGORY INTERCEPT POSITIVE, NEGATIVE RESPONSES	• Encode target word • Select and execute response
JUOLA CATEGORY SLOPE POSITIVE, NEGATIVE RESPONSES	• Retrieve category from LTM • Compare categories (serial)
CLARK AND CHASE "BASE" TIME	• Encode sentence • Encode picture • Construct kernel representations • Select and execute response
CLARK AND CHASE "BELOW" TIME	• Transform "below" representation
CLARK AND CHASE "NEGATION" TIME	• Transform "negation" representation
CLARK AND CHASE "COMPARISONS" TIME	• Transform truth indices
COLLINS AND QUILLIAN SUPERSET INTERCEPT	• Encode sentence • Construct kernel representation • Select and execute response
COLLINS AND QUILLIAN SUPERSET SLOPE	• Retrieve superset information from LTM
COLLINS AND QUILLIAN PROPERTY INTERCEPT	• Encode sentence • Construct kernel representation • Retrieve property information from LTM
COLLINS AND QUILLIAN PROPERTY SLOPE	• Retrieve superset information from LTM
SHEPARD AND TEGHTSOONIAN LAG FUNCTION EXPONENT, INTERCEPT	• Encode numbers • Store items in LTM
SHEPARD AND TEGHTSOONIAN p ("hits"), PROPORTION CORRECT	• Retrieve numbers from LTM
SHEPARD AND TEGHTSOONIAN d'	• Judge strength of activation
SHEPARD AND TEGHTSOONIAN p ("false alarms"), β	• Select and execute response

NOTE: *The following three measures are presented in the results section but are not included here:*

1. Posner "different" which is based on calculations across the three conditions,
2. Meyer "phonemic facilitation" which indicates an individual subject's propensity towards phonemic or graphemic encoding, and
3. Baron SH/HN which indicates an individual subject's propensity towards acoustic or visual encoding.

measured across the set of variables; many operations are sampled more than once. Also, most variables measure more than one operation. These observations will be considered more fully below, when construct validity is discussed. Prior to that discussion, further data will be presented regarding the measurement properties of these variables -- namely, their reliabilities, practice effects, and descriptive statistics.

Test-retest reliabilities. Table 8 summarizes the range of test-retest reliabilities obtained for the set of 40 variables. It should be kept in mind that the test-retest correlations are not commonly used as a reliability criterion. More typically, split-half (or odd-even) correlations are reported. In the present case, these latter measures were not reported, but they are substantially higher than test-retest correlations. Also, these test-retest correlations are sensitive to "strategy" changes on the part of the subject. For example, it was suggested above (when reporting the group results) that the data indicate a probable shift in some subjects' approach to some of the tasks from one day to the next. This "strategy shift" rationalization of low test-retest correlations is particularly appealing for the Juola tasks where either of two different strategies (i.e., categorize the target stimulus before comparison with the memory set or perform this categorization for each target item-memory set pair) would enable subjects to perform the task. It is also appealing for the Clark and Chase task, where some of the transformations might become unnecessary for some subjects after a day of practice.

Practice. Table 9 summarizes the effect of practice on the set of 40 variables in terms of those showing significant ($p \leq .05$) and nonsignificant t values when mean RTs for Day 1 and Day 2 are compared. For those variables showing significant practice effects, most have high test-retest reliability. The exceptions to this rule are some of the Juola measures. Again, it can be argued that since different strategies could be used in this task, some subjects changed strategies from day to day; if those sub-

Table 8

Test-Retest Reliabilities

$.70 \leq r$

Posner category	Baron Homophone-Nonsense
Posner "different"	Clark and Chase "negation"
Baron Sense-Nonsense	Collins and Quillian intercept property
Baron Homophone-Sense	

$.50 \leq r \leq .69$

Posner physical	Juola category intercept positive
Posner name	Clark and Chase "base"
Posner category minus name	Collins and Quillian intercept superset
Meyer "word"	Shepard proportion correct
Meyer "nonword"	Shepard p ("hits")
Sternberg slope positive	Shepard p ("false alarms")
Sternberg intercept positive	Shepard d'
Sternberg intercept negative	

$r \leq .49$

Posner name minus physical	Juola category slope negative
Meyer encoding facilitation	Juola category intercept negative
Baron SH/HN	Clark and Chase "below"
Sternberg slope negative	Clark and Chase "comparisons"
Juola word slope positive	Collins and Quillian slope superset
Juola word intercept positive	Collins and Quillian slope property
Juola word slope negative	Shepard lag-exponent
Juola word intercept negative	Shepard lag-intercept
Juola category slope positive	Shepard β

Table 9

Measures Showing Significant and Nonsignificant Practice Effects $(p \leq .05)$

Significant Effects	Nonsignificant Effects
Posner physical	Posner category minus name
Posner name	Meyer encoding facilitation
Posner category	Baron Sense-Nonsense
Posner "different"	Baron SH/HN
Posner name minus physical	Sternberg intercept positive
Meyer "word"	Sternberg slope negative
Meyer "nonword"	Juola word slope positive
Baron Sense-Homophone	Juola word slope negative
Baron Homophone-Nonsense	Juola category intercept positive
Sternberg slope positive	Juola category intercept negative
Sternberg intercept negative	Clark and Chase "below"
Juola word intercept positive	Clark and Chase "comparisons"
Juola word intercept negative	Collins and Quillian slope superset
Juola category slope positive	Collins and Quillian intercept superset
Juola category slope negative	Collins and Quillian slope property
Clark and Chase "negation"	Collins and Quillian intercept property
Clark and Chase "base"	Shepard proportion correct
Shepard lag - intercept	Shepard lag - exponent
	Shepard p ("hits")
	Shepard p ("false alarms")
	Shepard d'
	Shepard β

jects who used an efficient strategy maintained it while those who were inefficient changed, a significant practice effect would be obtained.

Most of the variables with non-significant practice effects are "derived" scores, rather than "baseline" measures. For example, the Baron "Sense-Homophone" and "Homophone-Nonsense" baseline scores both showed significant practice effects; however, the ratio of these two measures, Baron "SH/HN", did not change.

Descriptive statistics. Table 10 presents detailed descriptive information for each of the 40 measures. This information includes the number of subjects in the sample, the mean, median, standard deviation, maximum, minimum, and test-retest correlation (reliability) for each variable. Also shown are the obtained frequency polygons -- the distribution of scores within the sample population. These data may prove useful for two purposes: first, as comparison data for future paradigmatic replications, and second, as the basis for an information-processing data base. The establishment of such a data base would be invaluable for the eventual design of assessment instruments; in addition, it might prove to be of substantial impact for the design of tasks or the assessment of personnel requirements in man-machine systems.

Construct Validity

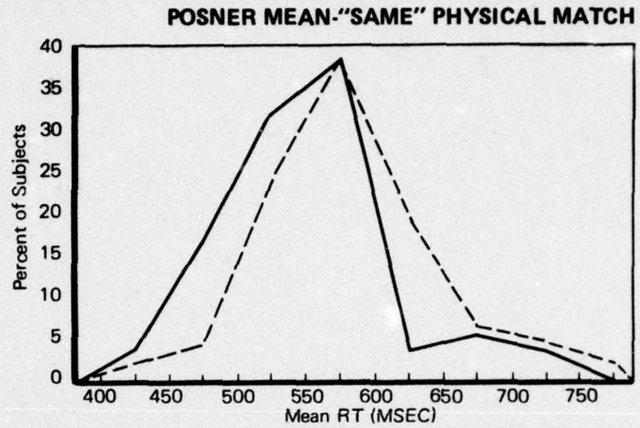
Overview. The concepts of converging operations and construct validity are relatively new in experimental psychology. While these concepts are of critical importance to any research concerned with test development, individual differences, and performance assessment, there are no formal experimental designs or statistical procedures extant for their evaluation. This is particularly true in the context of information-processing research, where the processes and operations are theorized to be "real" in the sense that they are discrete and take measurable amounts of time. Standard statistical procedures (e.g., factor analysis) are of marginal utility in

Table 10

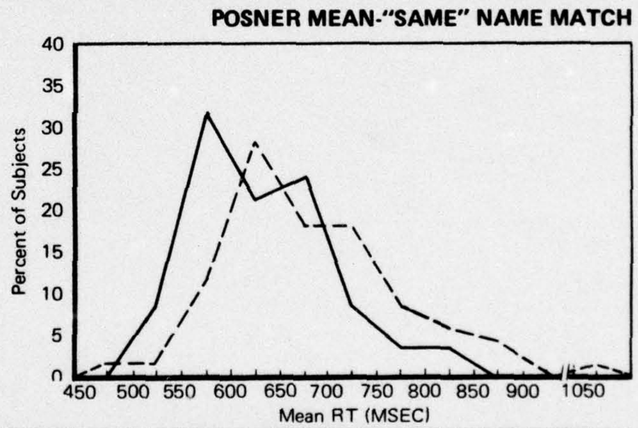
Descriptive Measures and Frequency Polygons

---- Day 1 —— Day 2

	Day 1	Day 2
N	50	37
Mean	585	547
Median	528	550
Standard Deviation	64	57
Maximum	794	721
Minimum	443	444
Reliability	.57	



	Day 1	Day 2
N	50	37
Mean	684	629
Median	669	625
Standard Deviation	100	71
Maximum	1056	830
Minimum	485	505
Reliability	.58	



	Day 1	Day 2
N	50	37
Mean	849	771
Median	812	773
Standard Deviation	173	121
Maximum	1425	1064
Minimum	594	575
Reliability	.78	

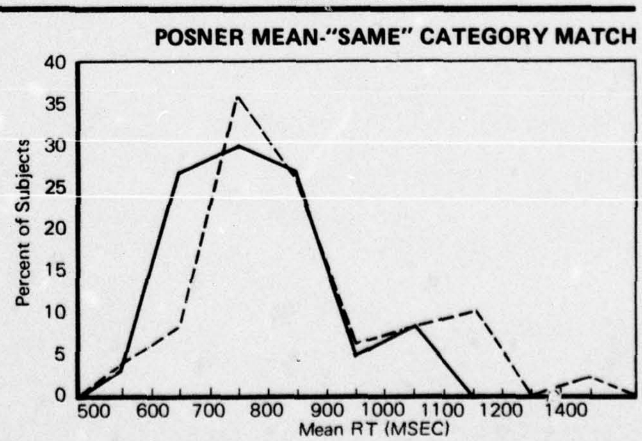
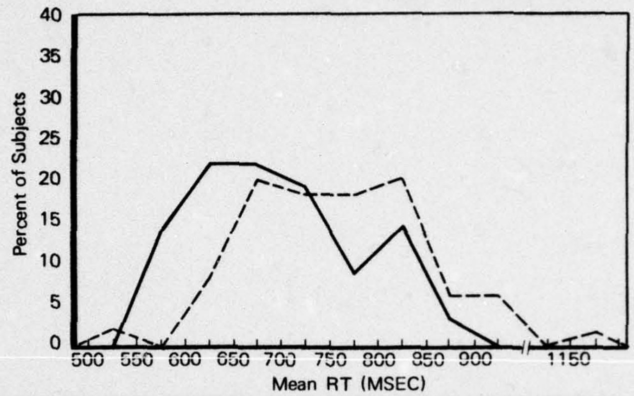


Table 10 (continued)

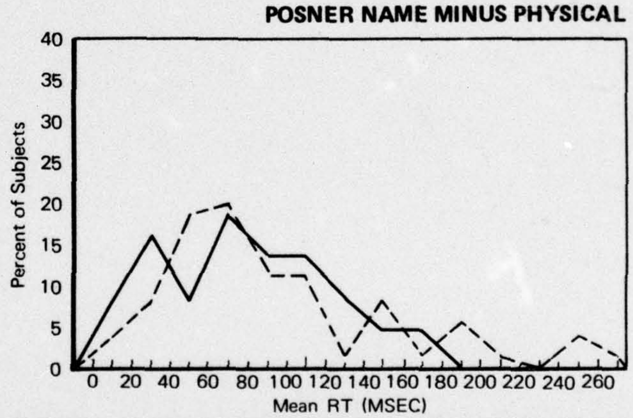
POSNER "DIFFERENT"

	Day 1	Day 2
N	50	37
Mean	761	693
Median	754	688
Standard Deviation	104	83
Maximum	1198	887
Minimum	532	553
Reliability	.81	



POSNER NAME MINUS PHYSICAL

	Day 1	Day 2
N	50	37
Mean	99	81
Median	80	80
Standard Deviation	62	45
Maximum	262	174
Minimum	9	12
Reliability	.29	



POSNER CATEGORY MINUS NAME

	Day 1	Day 2
N	50	37
Mean	164	137
Median	155	127
Standard Deviation	131	102
Maximum	469	347
Minimum	–31	9
Reliability	.69	

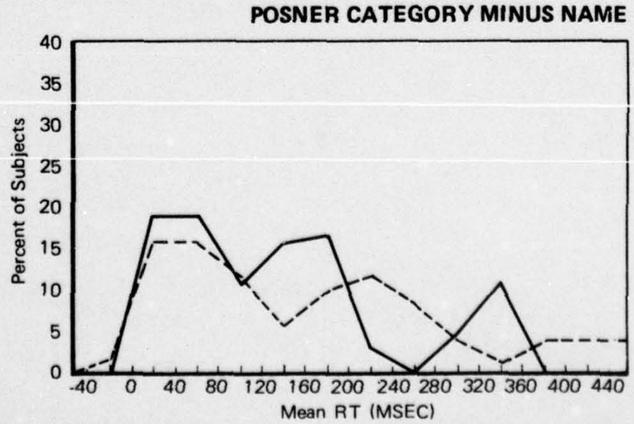
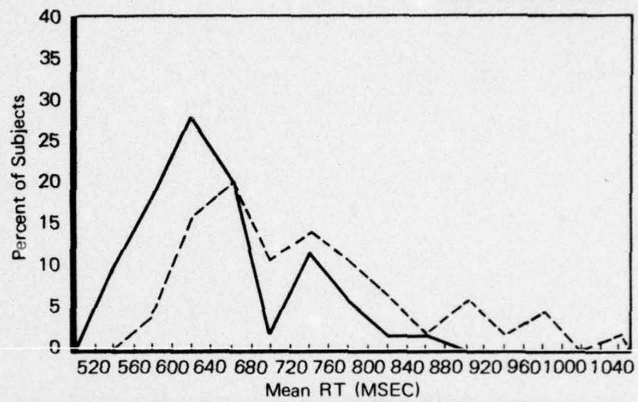


Table 10 (continued)

--- Day 1 —— Day 2

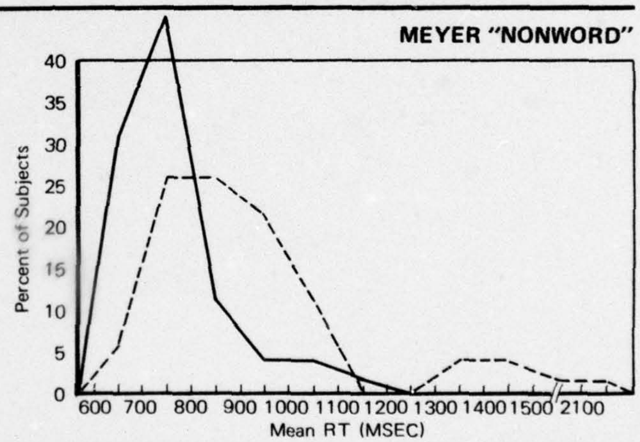
MEYER "WORD"

	Day 1	Day 2
N	54	50
Mean	736	647
Median	715	634
Standard Deviation	112	74
Maximum	1123	869
Minimum	582	522
Reliability	.66	



MEYER "NONWORD"

	Day 1	Day 2
N	54	50
Mean	916	756
Median	848	737
Standard Deviation	252	113
Maximum	2146	1195
Minimum	679	603
Reliability	.53	



MEYER ENCODING

	Day 1	Day 2
N	54	50
Mean	975	958
Median	978	952
Standard Deviation	78	75
Maximum	1153	1161
Minimum	775	794
Reliability	.42	

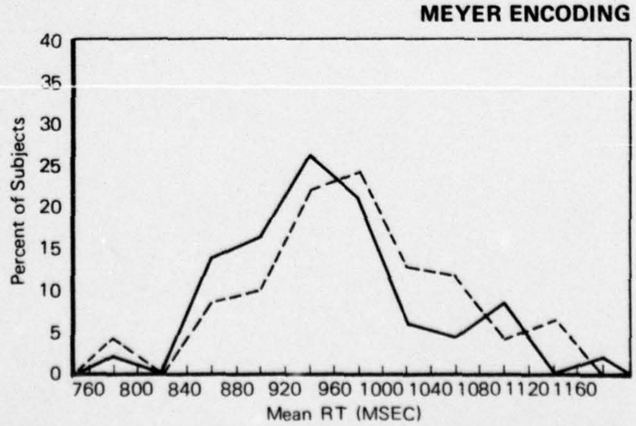
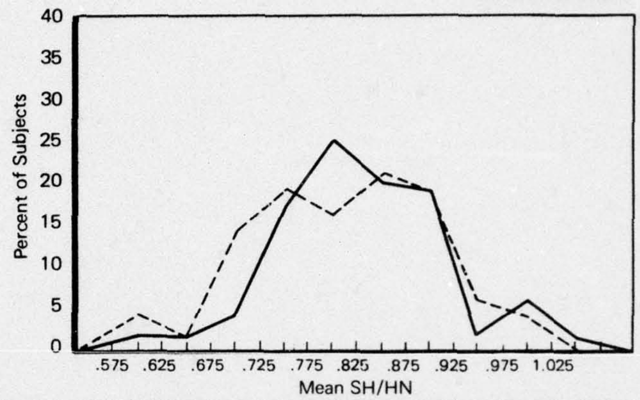


Table 10 (continued)

--- Day 1 —— Day 2

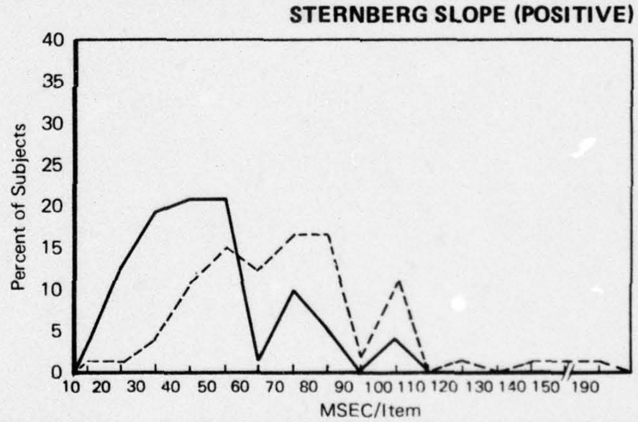
BARON SH/HN

	Day 1	Day 2
N	54	48
Mean	.81	.83
Median	.82	.83
Standard Deviation	.09	.09
Maximum	1.01	1.04
Minimum	.61	.62
Reliability	.37	



STERNBERG SLOPE (POSITIVE)

	Day 1	Day 2
N	54	48
Mean	75	49
Median	73	48
Standard Deviation	32	21
Maximum	190	107
Minimum	15	10
Reliability	.60	



STERNBERG INTERCEPT (POSITIVE)

	Day 1	Day 2
N	54	48
Mean	442	425
Median	431	413
Standard Deviation	88	78
Maximum	757	783
Minimum	261	312
Reliability	.52	

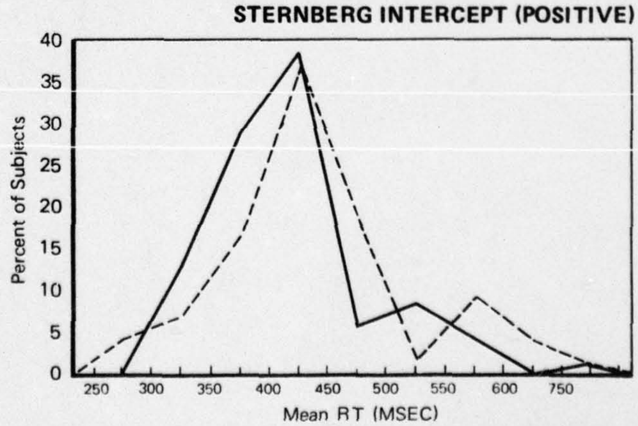
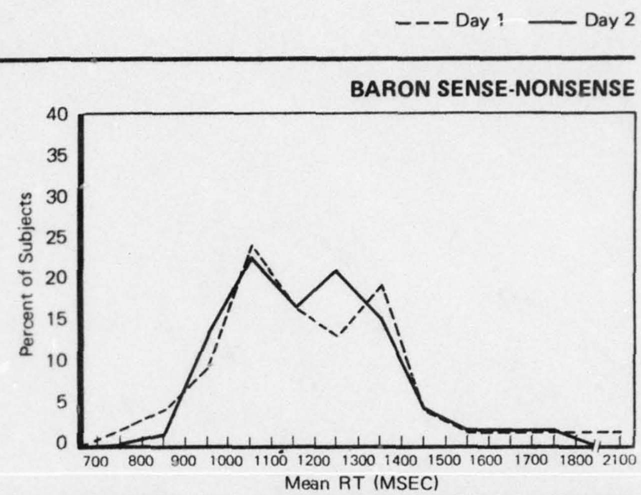


Table 10 (continued)

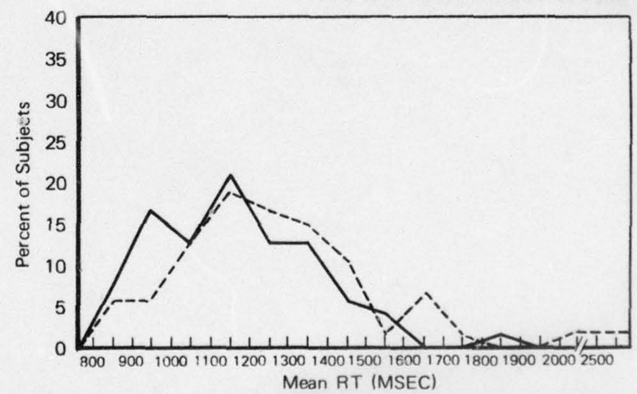
BARON SENSE-NONSENSE

	Day 1	Day 2
N	54	48
Mean	1205	1193
Median	1172	1181
Standard Deviation	246	197
Maximum	2113	1787
Minimum	760	830
Reliability	.83	



BARON HOMOPHONE-SENSE

	Day 1	Day 2
N	54	48
Mean	1289	1187
Median	1250	1165
Standard Deviation	300	241
Maximum	2590	1979
Minimum	851	852
Reliability	.90	



BARON HOMOPHONE-NONSENSE

	Day 1	Day 2
N	54	48
Mean	1579	1423
Median	1558	1450
Standard Deviation	306	235
Maximum	2616	2005
Minimum	1093	1027
Reliability	.47	

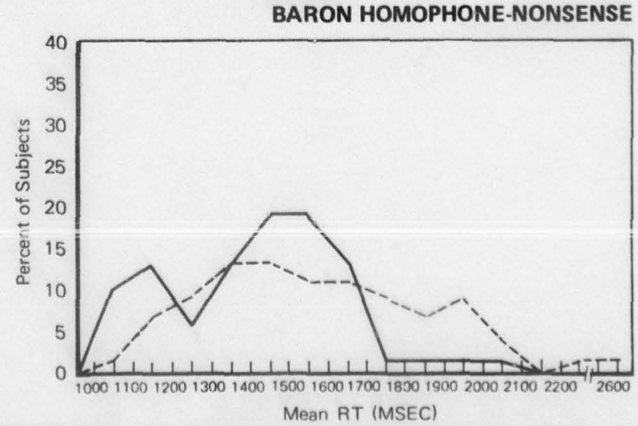
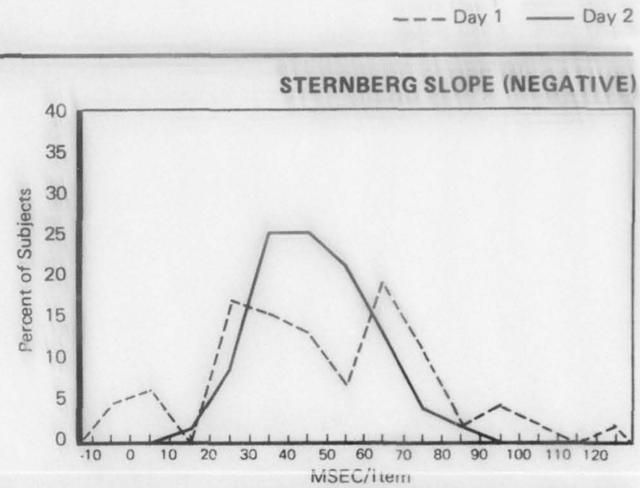
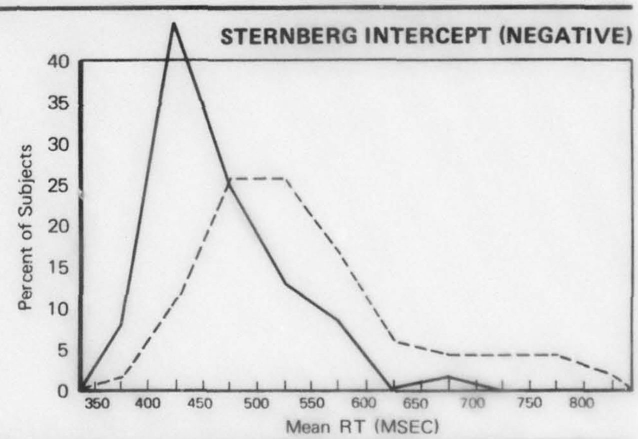


Table 10 (continued)

	Day 1	Day 2
N	54	48
Mean	48	47
Median	48	46
Standard Deviation	28	15
Maximum	121	88
Minimum	–17	12
Reliability	.45	



	Day 1	Day 2
N	54	48
Mean	536	464
Median	523	449
Standard Deviation	98	59
Maximum	847	654
Minimum	380	363
Reliability	.51	



	Day 1	Day 2
N	52	46
Mean	56	52
Median	54	52
Standard Deviation	32	24
Maximum	127	176
Minimum	–8	–12
Reliability	.19	

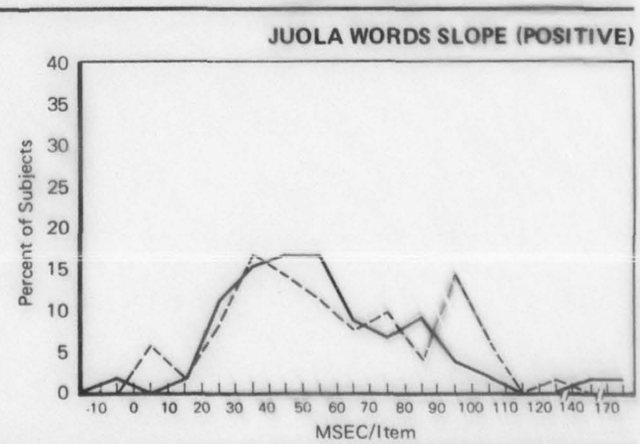
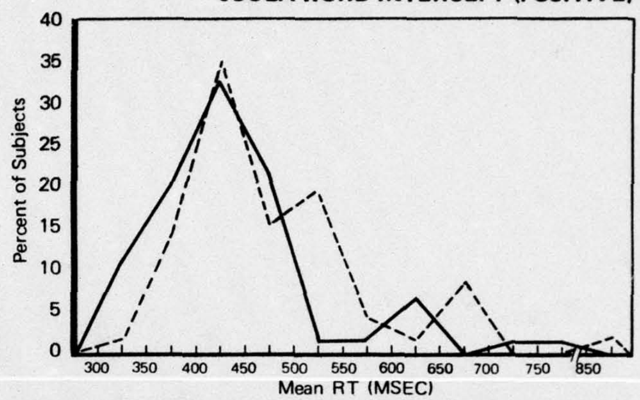


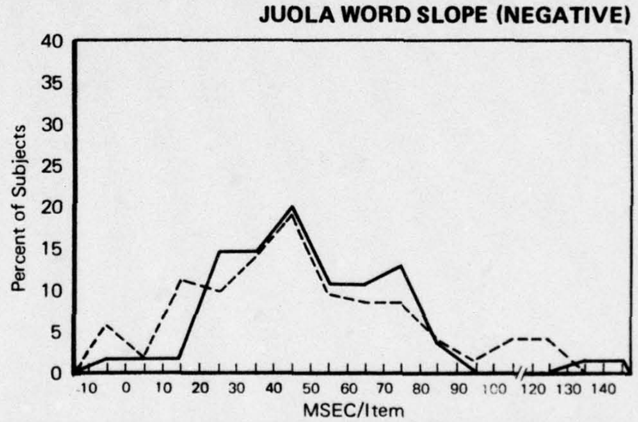
Table 10 (continued)

--- Day 1 —— Day 2

	Day 1	Day 2
N	52	46
Mean	483	446
Median	453	432
Standard Deviation	102	89
Maximum	876	779
Minimum	304	302
Reliability	.46	



	Day 1	Day 2
N	52	46
Mean	47	53
Median	45	47
Standard Deviation	32	31
Maximum	123	149
Minimum	−19	−16
Reliability	−.00	



	Day 1	Day 2
N	52	46
Mean	544	446
Median	498	470
Standard Deviation	145	67
Maximum	1148	766
Minimum	378	331
Reliability	.40	

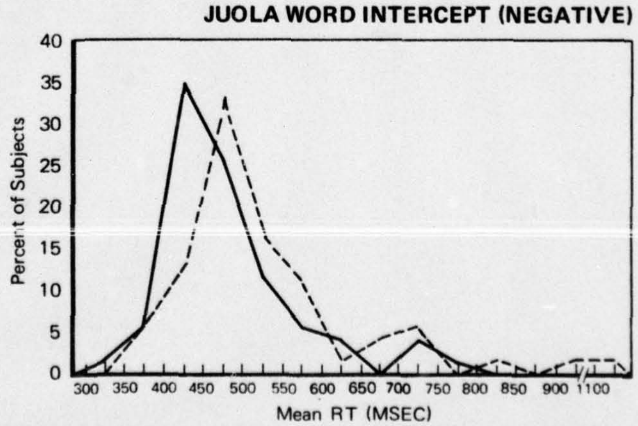
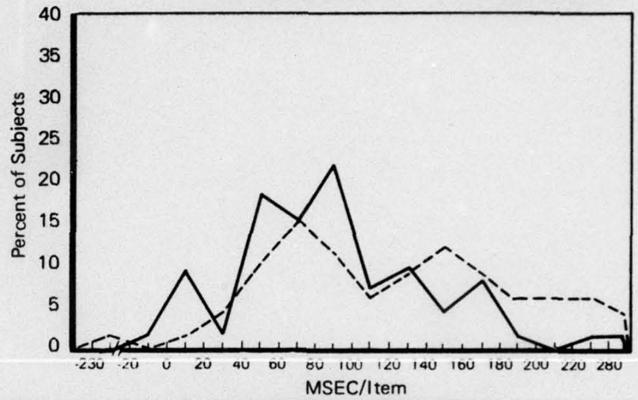


Table 10 (continued)

--- Day 1 —— Day 2

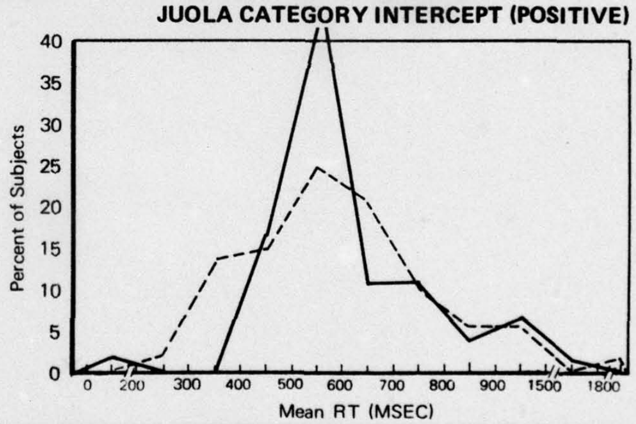
	Day 1	Day 2
N	52	46
Mean	122	93
Median	131	81
Standard Deviation	85	65
Maximum	287	297
Minimum	–239	–20
Reliability	.31	

JUOLA CATEGORY SLOPE (POSITIVE)



	Day 1	Day 2
N	52	46
Mean	611	637
Median	581	569
Standard Deviation	245	216
Maximum	1833	1582
Minimum	216	444
Reliability	.68	

JUOLA CATEGORY INTERCEPT (POSITIVE)



	Day 1	Day 2
N	52	46
Mean	214	140
Median	218	139
Standard Deviation	96	56
Maximum	530	310
Minimum	57	–3
Reliability	.32	

JUOLA CATEGORY SLOPE (NEGATIVE)

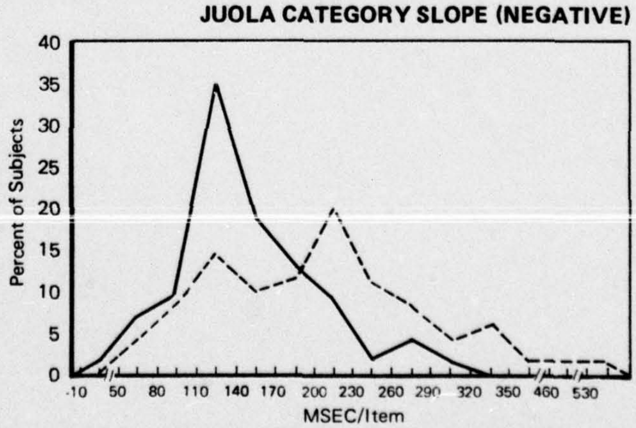
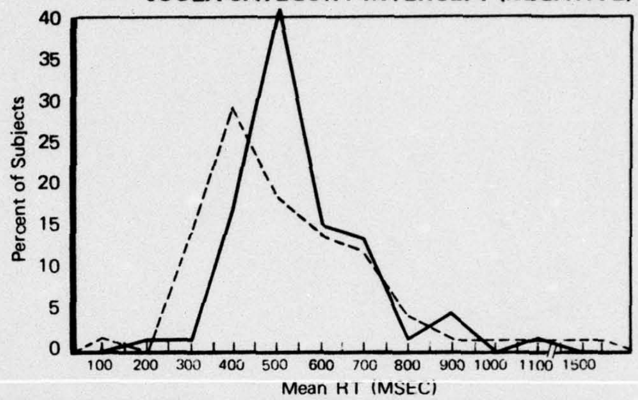


Table 10 (continued)

--- Day 1 —— Day 2

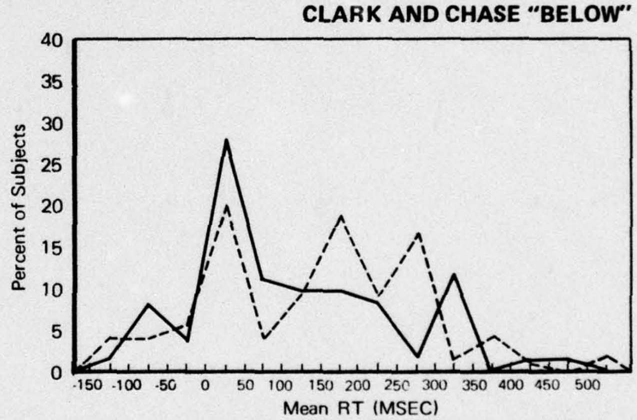
	Day 1	Day 2
N	52	46
Mean	575	595
Median	535	571
Standard Deviation	238	176
Maximum	1532	1109
Minimum	119	202
Reliability	.36	

JUOLA CATEGORY INTERCEPT (NEGATIVE)



	Day 1	Day 2
N	54	50
Mean	136	110
Median	163	88
Standard Deviation	155	149
Maximum	529	457
Minimum	−190	−151
Reliability	−.06	

CLARK AND CHASE "BELOW"



	Day 1	Day 2
N	54	50
Mean	829	685
Median	779	654
Standard Deviation	354	319
Maximum	2202	1467
Minimum	229	236
Reliability	.81	

CLARK AND CHASE "NEGATION"

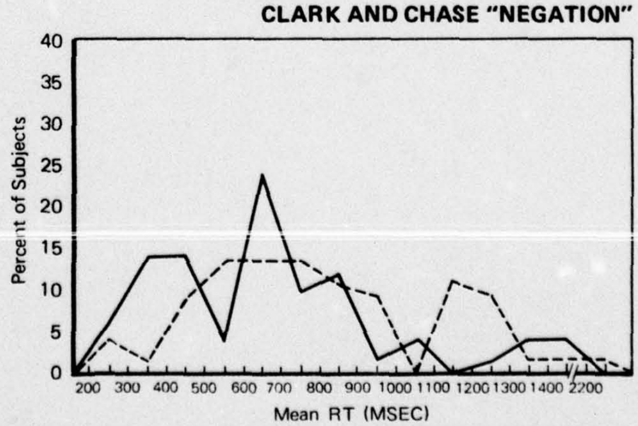
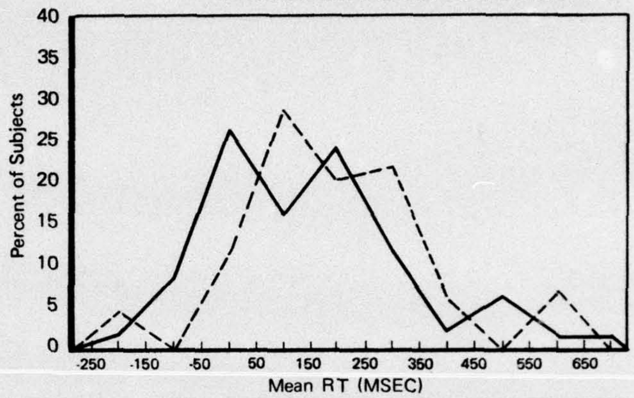


Table 10 (continued)

--- Day 1 —— Day 2

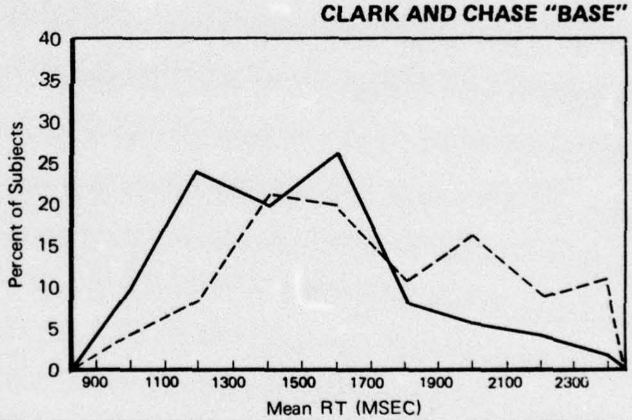
	Day 1	Day 2
N	54	50
Mean	200	146
Median	185	144
Standard Deviation	183	200
Maximum	648	726
Minimum	–286	–218
Reliability	.28	

CLARK AND CHASE "COMPARISONS"



	Day 1	Day 2
N	54	50
Mean	1735	1489
Median	1688	1450
Standard Deviation	404	330
Maximum	2475	2413
Minimum	1016	927
Reliability	.59	

CLARK AND CHASE "BASE"



	Day 1	Day 2
N	54	48
Mean	63	42
Median	67	48
Standard Deviation	57	77
Maximum	211	197
Minimum	–72	–247
Reliability	.21	

COLLINS AND QUILLIAN SLOPE (SUPERSET)

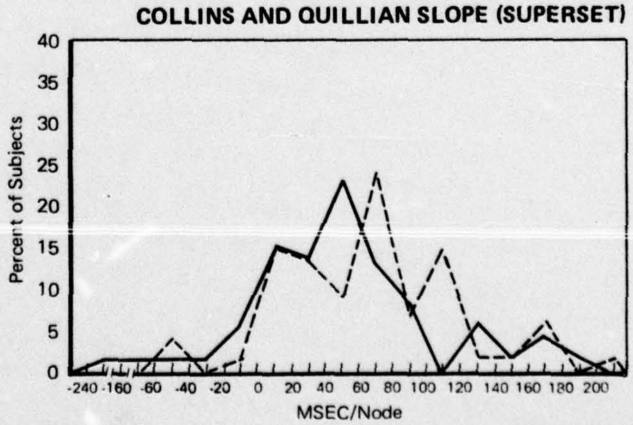
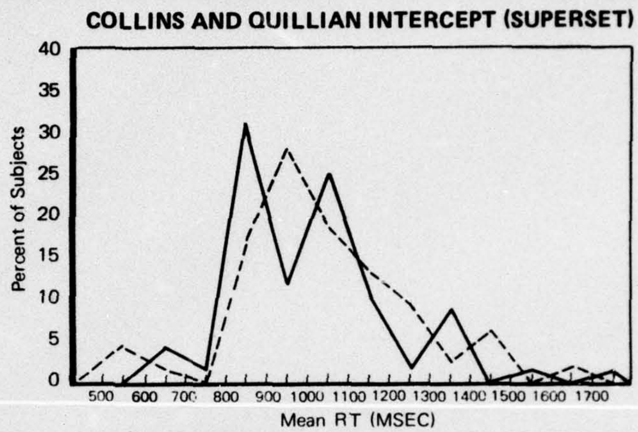


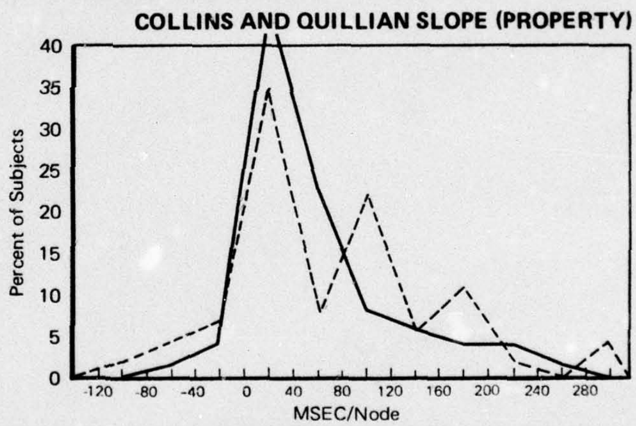
Table 10 (continued)

--- Day 1 —— Day 2

	Day 1	Day 2
N	54	48
Mean	1035	1017
Median	1005	1004
Standard Deviation	205	220
Maximum	1663	1786
Minimum	544	650
Reliability	.69	



	Day 1	Day 2
N	54	48
Mean	67	53
Median	50	38
Standard Deviation	89	75
Maximum	291	274
Minimum	–133	–83
Reliability	.16	



	Day 1	Day 2
N	54	48
Mean	1118	1121
Median	1081	1115
Standard Deviation	257	248
Maximum	1883	1911
Minimum	729	723
Reliability	.73	

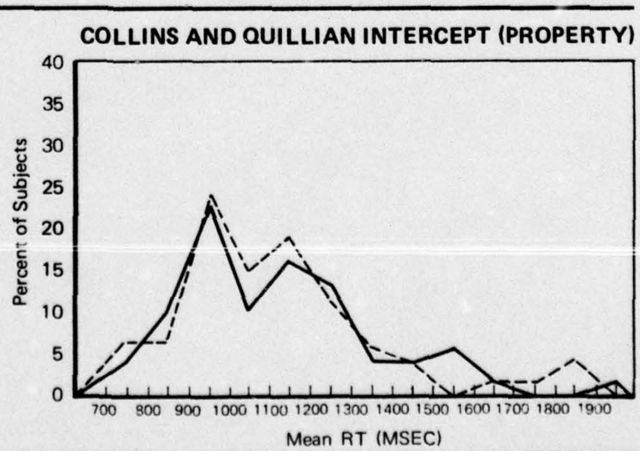
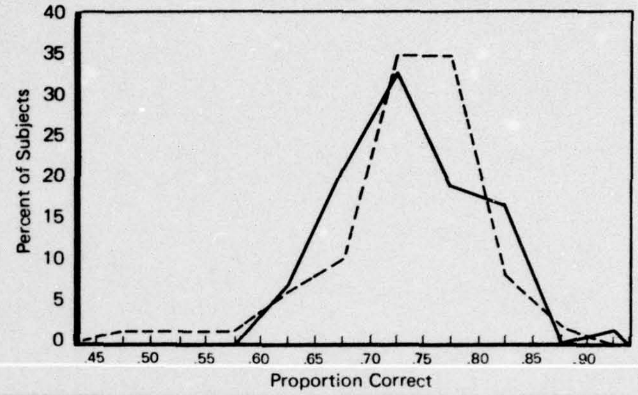


Table 10 (continued)

--- Day 1 —— Day 2

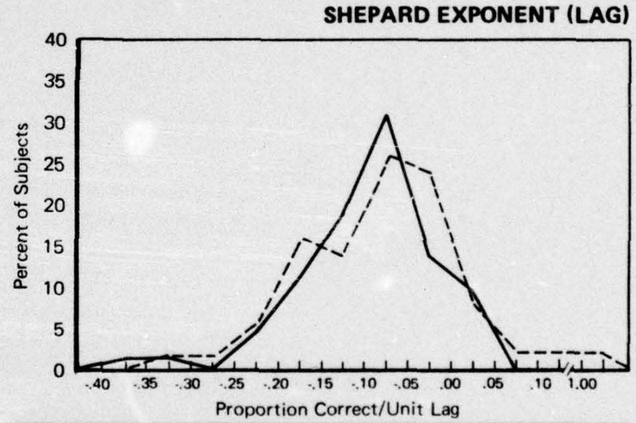
	Day 1	Day 2
N	52	42
Mean	73	73
Median	74	73
Standard Deviation	7	7
Maximum	86	93
Minimum	47	62
Reliability	.56	

SHEPARD PROPORTION CORRECT



	Day 1	Day 2
N	52	42
Mean	–.07	–.10
Median	–.08	–.08
Standard Deviation	.19	.08
Maximum	1.09	.03
Minimum	–.33	–.37
Reliability	.31	

SHEPARD EXPONENT (LAG)



	Day 1	Day 2
N	52	42
Mean	.86	.93
Median	.94	.91
Standard Deviation	.15	.12
Maximum	1.14	1.15
Minimum	.44	.62
Reliability	.21	

SHEPARD INTERCEPT (LAG)

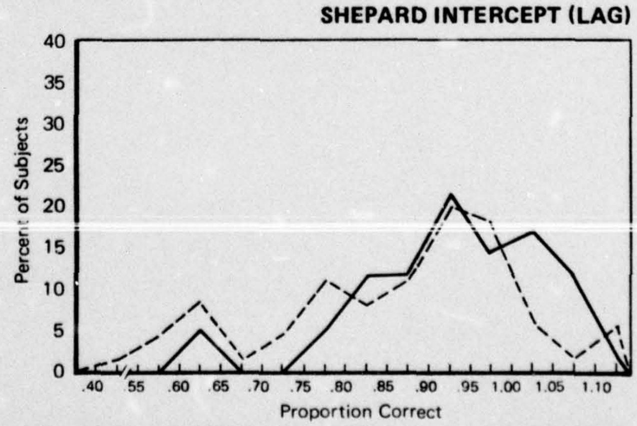
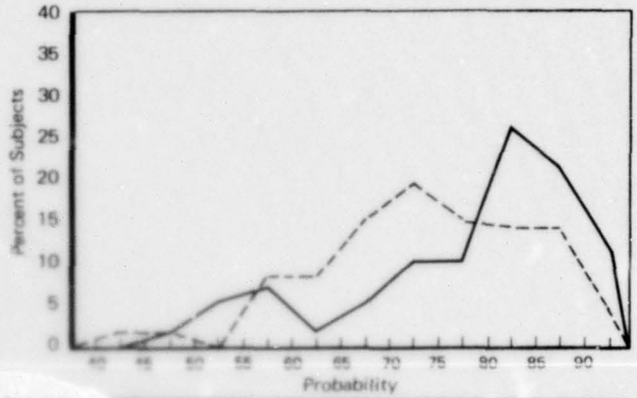


Table 10 (continued)

--- Day 1 —— Day 2

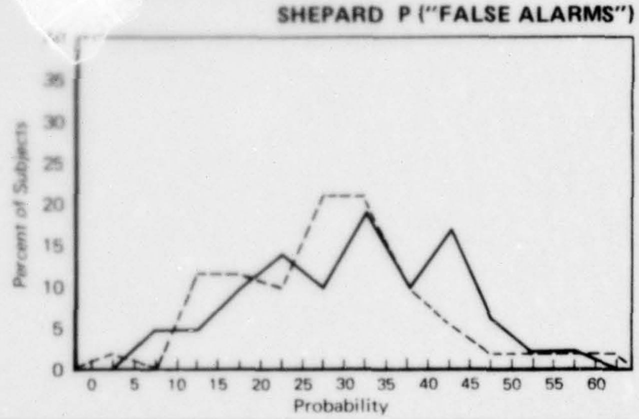
SHEPARD P ("HITS")

	Day 1	Day 2
N	52	42
Mean	73	77
Median	74	82
Standard Deviation	11	12
Maximum	94	92
Minimum	42	46
Reliability	.56	



SHEPARD P ("FALSE ALARMS")

	Day 1	Day 2
N	52	42
Mean	28	31
Median	29	32
Standard Deviation	12	12
Maximum	64	56
Minimum	4	6
Reliability	.67	



SHEPARD D'

	Day 1	Day 2
N	52	42
Mean	1.284	1.335
Median	1.317	1.238
Standard Deviation	.413	.452
Maximum	2.169	2.960
Minimum	.207	.614
Reliability	.62	

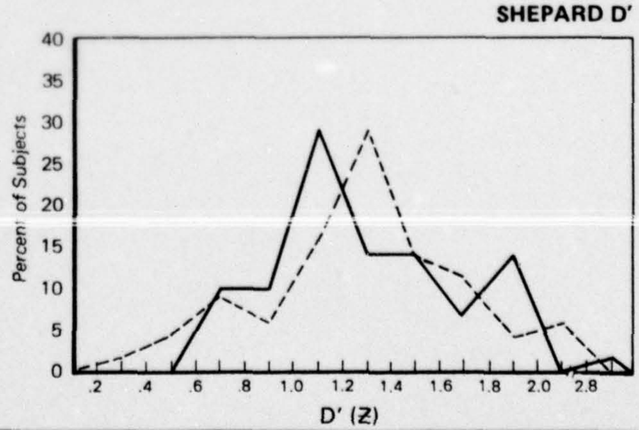
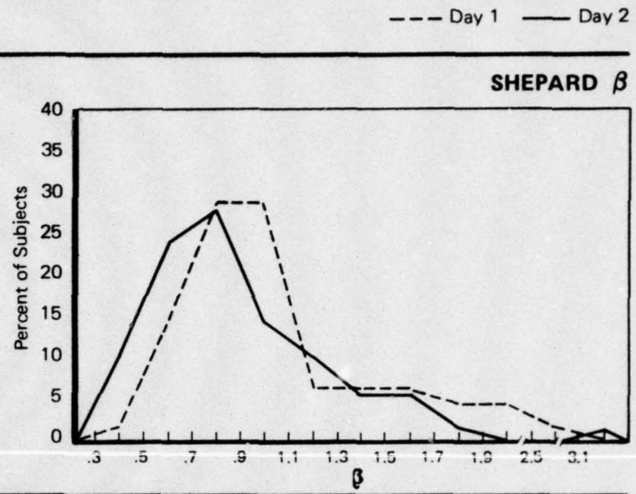


Table 10 (continued)

	Day 1	Day 2
N	52	42
Mean	1.086	.929
Median	.988	.825
Standard Deviation	.623	.501
Maximum	2.620	[illegible].200
Minimum	.353	.378
Reliability	.38	



that there is typically no presumption that the information-processing operations or their durations are independent or non-overlapping in time. Therefore, the analyses that were conducted in the present research can be viewed as "speculative": each analysis provides some information that can be interpreted as bearing on construct validity issues but has not been agreed upon by all researchers as the "correct" procedure.

The principal data used as inputs to the various analyses are the observed intra- and inter- task correlations shown in Table 11, the observed mean RTs for each of the variables, and a variable-by-operation matrix derived from Table 7. This matrix consists of the variables listed along one axis, the operations listed along the other, and the entries of "1" or "0" depending upon presence or absence of each operation in the composition of each variable. Using these inputs, three major analyses were conducted. The first was a general "model-fitting" procedure, using the correlations and the variable-by-operation matrix; the second was a stepwise regression analysis where the operations were used to predict the observed RTs; and the third was a multiple regression analysis, where estimates were obtained for the durations of the variables. Each of these analyses will be discussed below.

Model fitting. The notion of converging operations can be stated roughly in terms of experimental design -- the idea is to include in the same experiment some tasks that are hypothesized to involve a particular process and some tasks that do not. The pattern of empirical correlations among the tasks is then evaluated and inferences made about the validity of the particular process. In the present case, however, it is in practice impossible to interpret the empirical correlation matrices for several reasons. First, in the 40-by-40 matrix, there are 780 correlations; while it is conceptually possible to generate 780 hypotheses concerning the magnitude and direction of these correlations, it is simply not an efficient strategy to evaluate each hypothesis. Second, most of the variables measure more than one operation; it is far from apparent what the correlations between any two variables should be if (for example) they have one operation in common and a second operation

Table 11-A
Inter- and Intratask Correlations for Day 1*

		POSNER						MEYER			BARON				STERNBERG				JUOLA WORDS				JUOLA CATEGORIES				CLARK				COLLINS				SHEPARD					
		PHYSICAL MATCH	NAME MATCH	RULE MATCH	"DIFFERENT"	NAME MINUS PHYSICAL	RULE MINUS NAME	"WORD"	"NONWORD"	ENCODING FACILITATION	SENSE-NONSENSE	SENSE-HOMOPHONE	HOMOPHONE-NONSENSE	SH/HN	SLOPE POSITIVE	INTERCEPT POSITIVE	SLOPE NEGATIVE	INTERCEPT NEGATIVE	SLOPE POSITIVE	INTERCEPT POSITIVE	SLOPE NEGATIVE	INTERCEPT NEGATIVE	SLOPE POSITIVE	INTERCEPT POSITIVE	SLOPE NEGATIVE	INTERCEPT NEGATIVE	"BELOW"	"NEGATION"	"COMPARISON"	"BASE"	SUPERSET SLOPE	SUPERSET INTERCEPT	PROPERTY SLOPE	PROPERTY INTERCEPT	PROPORTION CORRECT	LAG-EXPONENT	LAG-INTERCEPT	PROBABILITY HITS	PROBABILITY FALSE ALARM	d'
POSNER	PHYSICAL MATCH																																							
	NAME MATCH	79																																						
	RULE MATCH	69	68																																					
	"DIFFERENT"	78	83	89																																				
	NAME MINUS PHYSICAL	28	81	40	55																																			
	RULE MINUS NAME	35	18	84	58	-06																																		
MEYER	"WORD"	46	59	58	61	49	35																																	
	"NONWORD"	45	64	47	55	56	17	64																																
	ENCODING FACIL.	-45	-43	-25	-36	-24	-05	-26	-33																															
BARON	SENSE-NONSENSE	67	70	69	74	47	40	69	59	-41																														
	SENSE-HOMOPHONE	64	70	62	71	48	32	63	48	-36	91																													
	HOMOPHONE-NONSENSE	63	73	75	77	55	46	66	62	-34	84	82																												
	SH/HN	26	19	05	15	04	-10	13	-05	-07	36	50	04																											
STERNBERG	SLOPE POSITIVE	29	32	51	56	23	43	38	37	02	33	25	41	-11																										
	INTERCEPT POSITIVE	41	39	24	26	21	05	00	10	-36	11	20	18	08	-34																									
	SLOPE NEGATIVE	30	47	45	48	45	24	25	38	-12	24	27	35	-01	42	19																								
	INTERCEPT NEGATIVE	53	39	50	55	11	39	31	23	-32	42	32	40	02	29	38	-22																							
JUOLA WORD	SLOPE POSITIVE	43	45	48	54	30	29	30	21	-13	43	39	34	24	37	08	30	32																						
	INTERCEPT POSITIVE	55	62	53	55	45	26	46	52	-29	49	50	52	03	24	44	31	37	-09																					
	SLOPE NEGATIVE	-02	07	08	03	12	05	03	03	-06	00	-01	02	-04	01	-05	18	-08	18	-02																				
	INTERCEPT NEGATIVE	63	62	63	70	37	38	45	40	-24	60	58	58	17	39	38	33	49	39	58	-50																			
JUOLA CAT.	SLOPE POSITIVE	15	33	25	24	38	08	20	08	17	22	35	29	23	15	03	21	-14	22	19	19	21																		
	INTERCEPT POSITIVE	31	20	23	34	02	18	23	30	-42	30	17	24	-07	20	07	04	42	07	24	-19	24	-73																	
	SLOPE NEGATIVE	16	30	33	26	31	23	34	18	07	26	20	32	01	27	-20	07	13	21	07	20	-01	26	06																
	INTERCEPT NEGATIVE	41	39	32	53	22	14	24	35	-38	48	50	38	22	29	25	32	29	29	36	-21	61	-08	42	-48															
CLARK	"BELOW"	18	01	04	11	-16	06	-04	-16	-08	08	07	-03	11	07	18	-12	33	06	09	-40	32	-24	38	-05	29														
	"NEGATION"	27	31	32	43	24	20	16	17	-23	15	23	19	06	34	14	37	13	26	19	06	22	11	20	14	25	17													
	"COMPARISON"	02	14	22	33	19	18	27	08	02	17	13	22	-07	52	-16	17	28	17	04	-05	18	-01	17	06	25	31	06												
	"BASE"	37	53	69	62	48	55	62	56	-18	59	53	72	-05	37	10	36	28	33	28	16	31	21	14	40	09	-23	33	07											
COLLINS	SUPERSET SLOPE	01	04	10	14	05	11	16	11	-04	25	20	06	08	02	-05	-03	12	17	11	-13	22	00	21	01	20	26	25	14	05										
	SUPERSET INTERCEPT	67	62	70	70	33	48	48	32	-37	58	54	65	18	38	34	28	60	45	33	12	42	02	32	25	28	11	21	22	60	-27									
	PROPERTY SLOPE	19	20	18	19	14	11	-04	06	03	01	06	14	-10	20	14	08	21	01	31	-24	38	17	06	20	08	22	15	-01	-03	01	19								
	PROPERTY INTERCEPT	60	55	64	64	29	45	62	37	-35	81	76	66	45	31	18	20	49	49	26	13	42	11	25	17	37	08	15	24	55	13	71	-21							
SHEPARD	PROPORTION CORRECT	-09	-06	-29	-22	00	-35	07	-05	-30	-10	-10	-20	03	-45	01	-09	-27	-22	-06	16	-26	-06	-01	-10	-06	-01	06	-17	-21	18	-28	-28	-05						
	LAG-EXPONENT	00	-08	-08	-04	-13	-05	-15	-09	-03	-06	00	-04	03	-22	08	11	-11	-15	-04	04	-06	-17	00	-15	00	-07	-16	00	-15	-02	-06	-11	-12	12					
	LAG-INTERCEPT	-23	-13	-25	-24	01	-25	-01	-17	-06	-10	-20	-21	-05	-15	-11	-31	-05	19	-33	24	-37	08	-17	12	-29	-09	05	-06	-06	07	-15	-21	-03	32	-45				
	PROBABILITY HITS	-27	-21	-24	-20	-08	-18	09	-22	09	-11	-12	-20	-03	-28	-16	-22	-11	-08	-24	33	-38	-03	-17	-02	-21	-07	02	01	-06	09	-19	-31	-02	58	27	42			
	PROB. FALSE ALARM	-15	-15	10	07	-10	24	-02	-17	42	-01	00	01	-02	26	-16	-11	21	19	-19	10	-07	02	-13	09	-11	-06	-03	20	18	-15	16	05	04	-63	12	03	25		
	d'	-07	03	-18	-18	11	-27	10	06	-24	-06	-09	-12	-10	-40	-02	-11	-21	-20	-02	18	-24	-03	-02	01	-16	-07	01	-25	-10	17	-23	-21	-03	92	07	29	54	-60	
	BETA	22	31	14	11	28	-03	07	58	-19	12	01	20	-21	16	07	23	-01	-01	29	-14	20	03	15	20	-01	-08	03	-15	15	08	-02	19	-05	-03	-27	-13	-63	-60	12

* Pearson product-moment correlations rounded to the nearest hundredth; decimals omitted. For df = 50, a value of r ≥ .32 is significant at the .01 (one-tailed) level.

Table 11-B
Inter- and Intratask Correlations for Day 2*

		POSNER						MEYER			BARON				STERNBERG				JUOLA WORDS				JUOLA CATEGORIES				CLARK				COLLINS				SHEPARD					
		PHYSICAL MATCH	NAME MATCH	RULE MATCH	"DIFFERENT"	NAME MINUS PHYSICAL	RULE MINUS NAME	"WORD"	"NONWORD"	ENCODING FACILITATION	SENSE-NONSENSE	SENSE-HOMOPHONE	HOMOPHONE-NONSENSE	SH/HN	SLOPE POSITIVE	INTERCEPT POSITIVE	SLOPE NEGATIVE	INTERCEPT NEGATIVE	SLOPE POSITIVE	INTERCEPT POSITIVE	SLOPE NEGATIVE	INTERCEPT NEGATIVE	SLOPE POSITIVE	INTERCEPT POSITIVE	SLOPE NEGATIVE	INTERCEPT NEGATIVE	"BELOW"	"NEGATION"	"COMPARISON"	"BASE"	SUPERSET SLOPE	SUPERSET INTERCEPT	PROPERTY SLOPE	PROPERTY INTERCEPT	PROPORTION CORRECT	LAG-EXPONENT	LAG-INTERCEPT	PROBABILITY HITS	PROBABILITY FALSE ALARM	d'
POSNER	PHYSICAL MATCH																																							
	NAME MATCH	71																																						
	RULE MATCH	54	52																																					
	"DIFFERENT"	71	66	84																																				
	NAME MINUS PHYSICAL	-05	67	17	19																																			
	RULE MINUS NAME	31	10	85	64	-19																																		
MEYER	"WORD"	41	62	57	53	44	37																																	
	"NONWORD"	48	60	54	62	34	33	73																																
	ENCODING FACIL.	24	06	17	01	-16	19	-18	-20																															
BARON	SENSE-NONSENSE	27	46	32	36	37	13	66	76	-44																														
	SENSE-HOMOPHONE	17	45	39	38	46	20	70	71	-33	87																													
	HOMOPHONE-NONSENSE	36	58	42	47	44	17	68	73	-41	85	75																												
	SH/HN	-09	-01	14	04	08	15	17	13	12	25	45	-12																											
STERNBERG	SLOPE POSITIVE	25	32	32	34	19	20	59	48	-26	40	39	45	-02																										
	INTERCEPT POSITIVE	40	49	61	54	28	46	29	26	34	14	24	16	23	-35																									
	SLOPE NEGATIVE	30	54	62	58	45	39	46	55	-15	31	35	38	01	43	36																								
	INTERCEPT NEGATIVE	53	55	76	72	22	61	52	47	17	38	52	41	36	20	71	38																							
JUOLA WORD	SLOPE POSITIVE	19	19	13	20	08	06	30	21	01	25	01	22	-18	34	-05	02	-07																						
	INTERCEPT POSITIVE	09	34	34	29	39	27	41	33	-10	26	52	18	34	05	45	31	57	-44																					
	SLOPE NEGATIVE	17	34	51	50	30	40	23	35	-19	32	39	45	13	51	14	56	57	-05	22																				
	INTERCEPT NEGATIVE	17	17	05	08	17	04	40	21	-09	07	23	-05	06	06	17	17	06	-05	54	-41																			
JUOLA CAT.	SLOPE POSITIVE	08	18	-03	02	17	-11	54	33	-37	59	47	51	06	39	-17	-21	08	42	13	-08	15																		
	INTERCEPT POSITIVE	31	33	41	45	14	31	10	22	-05	01	16	17	-06	28	16	52	41	-44	35	57	07	-46																	
	SLOPE NEGATIVE	19	02	08	16	-17	04	21	23	-27	34	33	35	-13	38	-22	01	04	21	01	13	05	49	-07																
	INTERCEPT NEGATIVE	36	56	41	51	41	27	39	43	-16	28	32	32	08	28	29	47	51	-22	50	45	23	-08	73	-35															
CLARK	"BELOW"	08	17	12	23	16	01	07	02	03	-10	01	00	-06	-15	28	14	05	09	-01	-16	29	-06	-07	00	-01														
	"NEGATION"	46	41	49	58	10	33	20	30	-07	10	26	38	-17	17	34	39	53	-34	42	46	01	-13	76	20	54	18													
	"COMPARISON"	24	11	23	28	-10	24	14	19	-09	-07	05	11	-09	21	-02	06	35	-29	23	26	07	-02	48	21	33	21	57												
	"BASE"	33	50	55	57	35	36	64	62	-21	61	57	63	11	67	16	63	48	37	12	60	03	33	14	30	24	-06	13	-11											
COLLINS	SUPERSET SLOPE	-33	-17	-36	-34	10	-38	10	-04	-52	16	07	00	10	11	-42	-18	-43	-01	-02	-37	23	37	-30	-15	01	00	-33	-24	-11										
	SUPERSET INTERCEPT	54	53	55	67	18	52	55	68	07	56	58	58	22	20	60	41	81	08	39	50	-03	13	27	17	33	-06	44	22	52	-54									
	PROPERTY SLOPE	-08	12	00	02	25	01	11	30	-16	17	21	34	-16	00	16	08	09	05	08	24	-14	19	-01	10	04	12	23	18	06	-07	20								
	PROPERTY INTERCEPT	25	40	29	32	29	09	61	52	-28	81	79	64	46	45	06	15	46	16	25	28	03	58	06	22	34	-07	12	02	54	22	44	-17							
SHEPARD	PROPORTION CORRECT	-01	07	-18	-12	10	-15	-06	-16	01	-10	-11	13	-33	-11	-09	-25	-09	-11	-09	-04	-15	09	05	-06	09	-14	10	09	-26	08	-18	38	-10						
	LAG-EXPONENT	07	10	-12	-02	07	-30	-06	08	00	-06	04	19	-22	-01	-15	-02	-05	-27	-15	02	-20	-04	06	15	-11	17	27	31	-09	-07	02	24	-03	36					
	LAG-INTERCEPT	00	06	-10	10	08	01	12	-15	-28	-01	02	-01	-03	-07	08	-05	05	11	14	-14	39	12	04	-12	23	10	00	01	07	20	-13	02	08	29	-26				
	PROBABILITY HITS	10	10	-22	05	04	-22	-04	-07	-12	-06	01	13	-21	-13	-07	-16	01	-15	-04	-07	00	04	05	01	08	13	22	25	-10	-02	00	24	03	60	65	50			
	PROB. FALSE ALARM	13	04	-03	19	-08	-07	04	13	-15	06	15	-01	16	-01	02	10	10	-02	06	-05	19	-05	-01	07	-01	29	10	16	19	-09	20	-19	17	-52	27	20	36		
	d'	-02	09	-15	-06	15	-14	-11	-18	00	-13	-16	12	-38	-14	-07	-22	-07	-06	-13	00	-19	04	03	-05	08	-08	07	10	-21	-01	-17	38	-15	97	39	32	66	-43	
	BETA	-08	-05	13	-09	01	15	-04	-08	13	-01	-15	-15	05	03	06	03	-08	08	00	00	-06	02	-03	-04	-02	-20	-20	-22	-10	07	-12	-13	-11	-08	-60	-35	-81	-76	-16

* Pearson product-moment correlations rounded to the nearest hundredth; decimals omitted. For df = 35, a value of r ≥ 38 is significant at the .01 (one-tailed) level.

not in common. Thirdly, there is no assumed independence of operations among tasks; that is, it is entirely possible that all measures are positively intercorrelated or that the correlations are mediated by higher-level "strategies." Therefore, an alternate procedure was developed.

This alternate procedure involved the calculation of the theoretical "distance" between each pair of variables in terms of the component operations. The variable-by-operation matrix was examined and pair-wise distances were calculated via the simple procedure of counting all operations present in both variables and dividing that sum by the total number of operations present in either task. (Other distance measures were tried, including the calculation of the Euclidian distance between tasks and a distance measure based on other assumptions about the zero-zero operation match; since all results were approximately equivalent, and the one described above is conceptually the easiest to understand, only this one procedure will be discussed.) From these calculations, a theoretical intervariable distance matrix was constructed. Finally, the correlation between the two matrices -- the empirical intercorrelations and the theoretical distances -- was calculated. This correlation was obtained separately for each day of testing and for two configurations of the variable-by-operation matrix. The first configuration included the operations of encoding, constructing, transforming, storing, retrieving names, retrieving categories, searching STM, searching LTM, comparing, and responding. The second configuration separated the various types of encoding: encoding of letters, letter strings, phrases, numbers, words, sentences, and pictures.

The results were as follows: For the reduced operations configuration (i.e., with a single encoding operation), a Pearson product-moment correlation of $r = .36$ was obtained on Day 1 and $r = .37$ was obtained on Day 2. For the complete set of operations (i.e., with the inclusion of seven types of encodings) the obtained values were $r = .22$ and $r = .24$ on Day 1 and Day 2 respectively. In addition, the Day 1-Day 2 "reliability" of the empirical correlation matrix was calculated; the obtained correlation was $r = .70$.

It is difficult to say whether or not these correlations represent "good" or "bad" model fits. Certainly, the fact that non-zero correlations were obtained can be interpreted positively, especially given the somewhat arbitrary original selection of operations. Moreover, it is clear that "improvements" of the fits could be accomplished if the variable-by-operation matrix were modified iteratively. Nevertheless, we believe that this fairly simple model-fitting procedure is a potentially valuable tool; in addition, we are encouraged by the positive relationship between the theoretical and empirical matrices.

Regression analyses. Another procedure that was used to "evaluate" the validity of the hypothetical operations was to consider the operations as predictors of the empirical measures in a regression paradigm. Basically, each obtained measure was considered as being composed of one or more operations that could be added together linearly to produce an observed response time. Since there were 35 such measures (the two Shepard and Teghtsoonian best-fit parameters were dropped, as were the three measures noted at the bottom of Table 7, since they cannot be interpreted as latency measures), there was a set of 35 simultaneous equations to solve. In the general linear models procedure, the model fit can be evaluated directly in terms of the obtained multiple R (the proportion of variance accounted for by the entire set of predictors); in addition, this procedure generates parameter estimates as beta weights (since the predictor matrix contained only ones and zeros) for each of the predictor variables. A stepwise regression procedure would reveal the predictor variables that account for the most variance.

Eight regression analyses were performed. Four of the eight were the general linear models procedure (using the complete variable-by-operation matrix and the reduced matrix for each of two days) and four were stepwise regressions (of the same type). Considering the general procedure first, the obtained R^2s were as follows: For Day 1, $R^2 = .72$; for Day 2, $R^2 = .70$. For the reduced operations matrix, these values were .71 and .70 for Day 1 and Day 2, respectively. All of these values were statistically significant ($F<.001$, df = 34). Likewise, the final R^2s of the stepwise regressions were

highly significant for all four analyses (the complete and reduced operations for Days 1 and 2), the obtained $R^2 = .66$ ($F<.001$, df = 34).

The pattern of parameter estimates for the general linear models procedure also supports the construct validity of the operations, in that the magnitudes of these estimates are intuitively in line with expectations. Some of these estimates are shown below. (It should be noted that an infinite number of solutions to the normal equations exists. Hence, some of the parameter estimates are biased in unknown directions. These biased parameters are followed by the letter B. Numbers are in msec.)

	Day 1 (All)	Day 2 (All)	Day 1 (Reduced)	Day 2 (Reduced)
Constructing	718	771	815	738
Transforming	269	184	311	209
Searching (LTM)	276B	260B	284B	179B
Searching (STM)	180	109	47	38
Comparing	135	127	196	161
Responding	341	259	401	289

In the stepwise regression analyses, constructing and responding accounted for significant components of the observed variance.

Conclusions. The results from several other analyses (as well as a more extensive presentation of the above analyses) could have been provided; however, the general pattern of results is consistent. The theoretical operations hypothesized to determine task performance do, both empirically and inductively, account satisfactorily for significant aspects of performance. Naturally, the definitions of information-processing constructs and the assignment of these constructs to variables should both be iterative activities. Likewise, the analysis procedures should be examined carefully and hopefully improved upon; it should be possible to develop some

standard construct validation procedures. However, to the extent that the present experiment has shed light on some of these issues, further research in the information-processing analysis of performance will benefit.

REFERENCES

Baron, J. Phonemic stage not necessary for reading. Quarterly Journal of Experimental Psychology, 1973, 25, 241-246.

Baron J. & McKillop, B. J. Individual differences in speed of phonemic analysis, visual analysis, and reading. Acta Psychologica, 1975, 39, 91-96.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "structure of intellect." Princeton, New Jersey: Educational Testing Service, May 1974.

Clark, H. H. & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Collins, A. M. & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

Juola, J. F. & Atkinson, R. C. Memory scanning for words versus categories. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 522-527.

Juola, J. F. & McDermott, D. A. Memory search for lexical and semantic information. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 567-575.

Melton, A. W. Individual differences and theoretical process variables: General comments on the conference. In R. M. Gagné (ed.) Learning and individual differences. Columbus, Ohio: Charles E. Merrill, 1967.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1969.

Posner, M. I. Abstraction and the process of recognition. In GH Bower & JT Spence (eds.) The psychology of learning and motivation: Advances in research and theory. New York: Academic Press, 1969.

Posner, M. I. & Mitchell, R. F. Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

Rose, A. M. Human information processing: An assessment and research battery. Human Performance Center Technical Report No. 46, Ann Arbor, Michigan, 1974.

Rose, A. M. & Fernandes, K. An information-processing approach to personnel management. Paper presented at the Office of Naval Research Contractors Conference on Job Analysis, Job Design, and Employment Criteria. December 1976.

Rubenstein, H., Garfield, L. & Millikan, J. A. Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 487-494.

Shepard, R. N. & Teghtsoonian, M. Retention of information under conditions approaching a steady state. Journal of Experimental Psychology, 1961, 62, 302-309.

Sternberg, S. Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 1975, 27, 1-32.

Sternberg, S. Two operations in character recognition: Some evidence from reaction-time measurements. Perception & Psychophysics. 1967, 2, 45-53.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. Acta Psychologica, 1969, 30, 276-315.

APPENDIX A

INSTRUCTIONS FOR EXPERIMENT I

Introduction

"The research you are about to take part in is one phase of a larger project designed to help understand basic human information processing capacities and limitations.

The results of this project will be used to improve educational and vocational guidance programs. The project will, for example contribute to the matching of individual qualifications and characteristics as needed for specific jobs and to the development of training programs for various occupations and professions.

Your participation in this project will require attendance at two sessions held on consecutive days, and consisting of approximately two hours each session.

During each session you will be asked to complete a series of simple tasks. These tasks involve simple matching and memory decisions, and do not test your knowledge of general information, your intelligence or your personality.

You should be able to complete the tasks within 2 hours. At the end of the second session you will be paid $20 for your participation. Are there any questions?

Read through the following description of the first task, and ask any questions you may have."

MEYER

"In this first task, you will see strings of letters. Your job will be to decide whether or not the letter strings spell a common English word. This is not a test of how many words you know; rather, we are measuring how quickly and accurately you can decide whether a string of letters is a word.

At the start of each trial, you will see two dots on the screen, one above the other. The appearance of these two dots will alert you that a letter string is about to appear. After a short delay, the top point will be replaced by a string of letters. You will decide whether or not it is a word. If it is, press the key on your left, as quickly as you can. If the letter string does not spell a word, press the key on your right, as quickly as you can. After you make your response, there will be a short delay and then the lower dot will be replaced by a second letter string. You will again judge whether or not this string is a word. If it is, you will again press the left key; if it is not, press the right key.

Each trial will consist of two letter strings and two responses. After each trial, the screen will provide information as to whether or not your responses were correct and how rapidly you made a correct response.

The first block of trials will be for practice. This will be followed by four more blocks, with a short rest between each.

Let us emphasize again that the English words used in this task are common, familiar words. Work as quickly and as accurately as you can. Any questions? If you are ready to begin, please press the button on your right."

CLARK and CHASE

"In this second task, we will be measuring how quickly and accurately you can determine whether a sentence is true or false. You will see a number of short sentences, each paired with a picture. The sentences will be on the left and the picture on the right. The picture shows either a star directly above a plus $\begin{smallmatrix}*\\+\end{smallmatrix}$ or a plus directly above a star $\begin{smallmatrix}+\\*\end{smallmatrix}$. The sentences claim to describe the pictures. For example, the sentences, "STAR IS ABOVE PLUS." and, "STAR ISN'T BELOW PLUS." are both true descriptions of the picture, $\begin{smallmatrix}*\\+\end{smallmatrix}$.

Your job is to read each sentence and to decide whether it is a true or a false description of the picture. If you think that the sentence describes the picture correctly, press the left hand key. If you think that the sentence does not give a true description of the picture, press the right hand key. You are to work as quickly as you can without making mistakes.

Each trial will begin with the appearance of a dot on the left side of the screen. Following a short delay, the dot will disappear and the sentence and picture will be shown. After you make your response, the screen will display the word "CORRECT" and your response time if you were correct or the word "WRONG" if you were incorrect.

The first list you will see is a practice list. This will be followed by five blocks of sentences with a short rest between blocks. Any questions?"

STERNBERG

"This third task measures how quickly and accurately you can recognize items that you have just seen. On each trial, you will see one or more numbers on the left side of the screen. You will be given a few seconds

to memorize the numbers, after which they will disappear. Following this study period, a single number will appear on the right side of the screen. You will decide whether or not this number was one of the numbers you have just memorized.

If the test number was present in the memory list, press the left key. If the test number was not part of the memory list, press the right key.

After each response, the screen will provide information as to whether or not your response was correct and how rapidly you made a correct response. The first block of trials will be for practice.

The practice block will be followed by five test blocks, with a short rest between each block. Please work as quickly as you can without making any errors. Any questions?"

COLLINS AND QUILLIAN

"This fourth task measures how quickly and accurately you can decide if a sentence is true or false. Each sentence is a factual statement like ZEBRAS HAVE STRIPES. If a sentence is generally true call it true and if it is generally false call it false. Do not waste time thinking of exceptions to generally true or generally false sentences; you will easily be able to tell the difference between the two.

Occasionally you will see a redundant sentence, such as A BOAT IS A BOAT. These sentences serve as a comparison for the other sentences. Treat them as true.

Each trial begins with a dot on the screen to warn you that a sentence is about to appear. When the sentence appears, respond by pressing the left key if the sentence is generally true; if the sentence is generally false, press the right key.

The first set of sentences will be for practice. This practice block will be followed by two test blocks with a short break between them. Following each response there will be a short delay, during which response time and error information will appear on the screen. Any questions?"

POSNER

Physical

"In this fifth task, you will be making a series of simple judgments. In the first condition we are concerned with how accurately and rapidly you can decide whether two displays are physically the same or different. You will see pairs of letters, one on the left side of the screen and one

on the right side. Your job is to judge whether the two letters are physically the same or if they are different. For example, the letter pair AA would be judged as the same, while the pair AB would be judged as different.

If they are the same, press the left key; if they are different, press the right key.

After you respond, the screen will display how fast you responded and whether or not you were correct. Please respond as rapidly as you can, without making errors.

Before the letters appear, two dots will be displayed. These dots will be your warning signal - a few seconds after the dots appear, the letters will come on. The first block of trials is for practice. Any questions?"

Name

"In this second condition you will again see pairs of letters. This time, you are to decide whether or not the letters have the same name. For example, the letter pair Aa would be judged as the same, while the pair Ab would be judged as different. If they are the same, respond by pressing the left key; if they have different names, press the right key. Work as quickly as you can without making any errors. The first block of trials is for practice. Any questions?"

Category

"In this third condition, you will again see pairs of letters. This time, you are to decide whether or not the two letters belong to the same category. If the two letters are both vowels or both consonants, respond by pressing the left key. For example, the letters AO are both vowels and should be judged as same; the letters TN are both consonants and would also be judged as same. If these letters are different - one is a vowel and the other is a consonant - press the right key. For example, the letter pair AB would be judged as different. Work as quickly as you can without making any errors. The first block of trials is for practice. Any questions?"

BARON

Sense-Nonsense

"The sixth task consists of three conditions. In each condition, you will see a series of short phrases. In this first condition your job will be to decide whether or not these phrases make sense. This task is designed to measure how quickly and accurately you can make this judgment; it is not a vocabulary test, nor is it a test of your ability to make sense out of nonsense phrases. Therefore, don't waste time trying to make sense out of

nonsense phrases; the difference between sense and nonsense will be obvious.

At the start of each trial, you will see a dot on the left side of the screen. The appearance of the dot will alert you that a phrase is about to appear. After a short delay, the dot will be replaced by the phrase. You will decide whether or not it makes sense. If it does, press the key on your left. If it does not, press the key on your right.

After each trial, the screen will display your response speed and accuracy. The first few trials will be for practice. These practice trials will be followed by two test blocks with a short rest between blocks. Any questions?"

Sense Homophone

"In this second condition, you will again see a series of phrases and you will again decide whether or not the phrases make sense. However, in this condition, some of the phrases have been constructed so that if you say them, they would sound as if they made sense. Therefore, you must be careful to judge the phrases as sense or nonsense on the basis of how they look and not on the basis of how they sound. As before, if a phrase makes sense, respond by pressing the left key; if it does not make sense, press the right key. The first few trials will be for practice. These practice trials will be followed by two test blocks. Again, work as quickly and as accurately as you can. Any questions?"

Nonsense Homophone

"In the next condition, you will again see a series of phrases and you will again decide whether or not the phrases make sense. However, in this condition all phrases will look like nonsense. You are to judge whether or not the phrases make sense by their sound. If a phrase sounds like it makes sense, press the left key; if it sound like nonsense, press the right key.

Again, work as quickly and as accurately as you can. The first few trials will be for practice, followed by two test blocks. Any questions?"

JUOLA

Words

"The seventh task consists of two parts. The first part measures how quickly and accurately you can recognize items that you have just seen.

On each trial, you will see from one to four words on the left side of the screen. You will be given a few seonds to memorize the words.

After this study period, a single word will appear on the right side of the screen. You will decide whether or not this word was one of the words you have just memorized.

If the test word was present in the memory list, press the left key. If the test word was not part of the memory list, press the right key.

After each response, the screen will provide information as to whether or not your response was correct and how rapidly you made a correct response.

The first block of trials will be for practice. The practice block will be followed by two test blocks, with a short rest between each block. Please work as quickly as you can without making any errors. Any questions?"

Category

"As you may have noticed, each of the words used in the first part of this task was the name of a category. The second part of this task is similar to the first in that you will again be shown from one to four category words on the left side of the screen for you to memorize.

Again, a single test word will appear on the right side of the screen. However, this time you will decide whether or not the test word is an example of one of the categories in the list you have just memorized. For instance, suppose that you see the words COLOR and RELATIVE on the left side for you to memorize. The test word might be BLUE. Since this is an example of the category COLOR, you would respond by pressing the left key. If the test word had been HORSE, you would respond by pressing the right key since HORSE is not an example of either COLOR or RELATIVE.

As before, the screen will provide information concerning your response speed and accuracy after each trial. The first block of trials will be for practice. This practice block will be followed by two test blocks with a short rest between them. Please work as quickly as you can without making any errors. Any questions?"

SHEPARD

"This eighth task is a little different from the rest of the tasks. This is a test of how well you can recognize numbers that you have previously seen.

You will be shown a series of three-digit numbers. For each three-digit number you must decide whether or not you have seen that number before in the series. If you have seen it, respond by pressing the left key. If you have not seen that number before, press the right key.

To illustrate, look at the following sequence of numbers and the correct responses to each:"

Number	Response
100	New (right key)
200	New (right key)
100	Old (left key)
300	New (right key)
200	Old (left key)

"The first number is, of course, always "New". This task does not measure how quickly you respond; however, there is a 10-second time limit for each response. If you take longer than 10 seconds, a "TOO SLOW" message will appear on the screen and the next number in the list will be shown. Therefore, you will have to work fairly rapidly. Within these constraints, try to work as accurately as you can.

In order to minimize confusion, there will not be any practice for this task. Therefore, if you have any questions, ask them now."